Near-Term Missions Standards Recommendations

December 2, 2002

SEEDS Near-Term Mission Standard Study Team

Contributors

Richard Ullman, NASA/GSFC, Study Team Lead
Jingli Yang, ERT, Study Team
Cheryl Craig, NCAR, Study Team
John Evans, GST, Study Team
Larry Klein, L-3 Analytics, Study Team
Dorian Shuford, ERT, Study Team
Siri Jodha Singh Khalsa, L-3 Analytics, Study Team
Matt Smith, UAH, Study Team

Table of Contents

<u>1.0</u>	<u>INTRODUCTION</u>	1
1.1	SEEDS GOALS AND STRATEGY	1
1.2	THE RATIONALE FOR STANDARDS	2
<u>1.3</u>	Assumptions	4
<u>1.4</u>	METHODOLOGY	4
<u>2.0</u>	NEAR-TERM MISSION AND HERITAGE MISSION STANDARDS	6
2.1	SEEDS NEAR-TERM MISSIONS	6
2.2	HERITAGE MISSION STANDARDS	7
<u>3.0</u>	LESSONS LEARNED.	. 12
3.1	LESSONS LEARNED ON IMPLEMENTING AND USING NASA EOS STANDARDS	. 12
	<u>.1.1 Landsat 7</u>	
_	<u>.1.2</u> <u>TERRA</u>	
	<u>.1.3</u> <u>AQUA</u>	
	. <u>1.4</u> <u>AURA</u>	
	.1.5 QuikSCAT/SeaWinds	
	.1.6 ACRIM	
	. <u>1.7 SeaWiFS</u>	
	<u>.1.8</u> <u>Jason-1</u>	
	LESSONS LEARNED ON IMPLEMENTING AND USING OTHER STANDARDS	
	.2.1 NOAA Standards	
_	.2.2 The Spatial Data Transfer Standard (SDTS)	
4.0	ESSENTIAL STANDARDS CONCEPTS	
4.1	A Comparison with Private, Ad-hoc, Binary Information Transfer	
4.2	MANDATORY VS. OPTIONAL ELEMENTS, PROFILES AND EXTENSIONS	
4.3	ABSTRACT VS. IMPLEMENTATION STANDARDS.	
4.4	CONTENT AND FORMAT VS. BEHAVIOR AND INTERFACE	
4.5	WEB-BASED DATA SERVICE STANDARDS	
<u>5.0</u>	STANDARDS EVALUATION	. 35
5.1	EVALUATION CRITERIA	. 35
5.2	Data Standards Evaluation	
5.3 5.4 5.5	METADATA AND DOCUMENTATION STANDARDS EVALUATION	. 45
<u>5.4</u>	CATALOG INTERFACE STANDARDS EVALUATION	. 47
<u>5.5</u>	<u>User Surveys</u>	
	.5.1 <u>Data Format Standards</u>	
<u>5</u>	.5.2 <u>Metadata Format Standards</u>	. 54
<u>6.0</u>	<u>SUMMARY</u>	. 57
<u>7.0</u>	<u>CONCLUSIONS</u>	. 60
<u>7.1</u>	Data Packaging Standards	. 60

NTMS recommends the following for packaging of near-term mission s	<u>tandard</u>
products.	60
7.1.1 Data Distribution Formats Recommendations	
7.1.2 Data Interchange Formats Recommendations	61
7.2 METADATA STANDARDS RECOMMENDATIONS	62
7.3 DOCUMENTATION STANDARDS RECOMMENDATIONS	63
7.4 CATALOG INTERFACE STANDARDS RECOMMENDATIONS	64
7.5 STANDARD EVOLUTION PROCESS & OTHER ACTIVITIES RECOMMENT	<u>DATIONS</u> 64
ACRONYM LIST	66

List of Tables and Figures

FIGURE 1.1.1 SIMPLIFIED ESE NETWORK DATA FLOW.	2
TABLE 2.1.1 SEEDS NEAR-TERM MISSIONS.	
TABLE 2.1.2 SEEDS NEAR-TERM MISSION STANDARDS	6
TABLE 2.2.3 SEEDS HERITAGE MISSIONS DATA MANAGEMENT INFORMATION	7
TABLE 4.4.1 VIEWPOINTS AND LEVELS OF ABSTRACTION	29
TABLE 4.4.2 CRITERIA FOR FORMAT STANDARDS	29
TABLE 4.4.3 CRITERIA FOR INTERFACE STANDARDS	30
TABLE 4.4.4 DATA MODELS AND SOFTWARE ACCESS LIBRARIES	
TABLE 5.2.1 DATA STANDARDS INTEROPERABILITY	
TABLE 5.2.2 DATA STANDARDS AVAILABILITY	
TABLE 5.2.3 DATA STANDARDS PORTABILITY.	
TABLE 5.2.4 DATA STANDARDS EVOLVABILITY	
TABLE 5.2.5 DATA STANDARDS EXTENSIBILITY	41
TABLE 5.2.6 DATA STANDARDS SELF-DESCRIBING	41
TABLE 5.2.7 DATA STANDARDS TOOLS SUPPORT	
TABLE 5.2.8 SEMANTIC COMPLETENESS	
TABLE 5.2.9 DATA STANDARDS EVALUATION	
TABLE 5.3.1 METADATA AND DOCUMENTATION STANDARDS EVALUATION	
TABLE 5.4.1 CATALOG INTERFACE STANDARDS EVALUATION	
TABLE 5.5.1 SURVEY RATINGS OF ATTRIBUTE IMPORTANCE.	51
TABLE 5.5.2 DATA STANDARDS SURVEY EVALUATION	
TABLE 5.5.3 SUMMARY OF SURVEY ESSAY QUESTIONS	
TABLE 5.5.4 METADATA STANDARDS SURVEY EVALUATION	
TABLE 5.5.5 SUMMARY OF METADATA SURVEY ESSAY QUESTIONS	

1.0 Introduction

1.1 SEEDS Goals and Strategy

SEEDS, previously called NewDISS, is the <u>Strategic Evolution</u> of the <u>Earth</u> Science Enterprise <u>Data Systems</u> meant to serve research and application needs in the next ten years. Its primary goal is to support NASA's Earth Science Enterprise (ESE), which, in turn, contributes to the US Global Change Research Program (USGCRP). As such, SEEDS is driven principally by the objectives of scientific research, but must also serve the needs of both scientific research and a wide variety of practical applications.

Future ESE data systems will consist of a heterogeneous mix of interdependent components derived from the contributions of numerous individuals and institutions. These widely varying participants will be responsible for data management functions including data acquisition and synthesis; access to data and services; and data stewardship.

"An important premise underlying the operation of [the ESE network of data systems and services] is that its various parts should have considerable freedom in the ways in which they implement their functions and capabilities. Implementation will not be centrally developed, nor will the pieces developed be centrally managed. However, every part of [the ESE network] should be configured in such a way that data and information can be readily transferred to any other. This will be achieved primarily through the adoption of common standards and practices [1]."

Figure 1.1.1 is a simplified data flow diagram of the ESE network of data systems and services [1]. Five types of data centers, namely Backbone Processing Centers, Principal Investigator (PI)-managed Mission Data Centers, Science Data Centers, Applications Data Centers, and Multimission Data Centers are shown in the diagram. Several data flows, such as data flows from PI-managed Mission Data Centers to Multimission Data Centers and vice versa, from Science Data Centers to Applications Data Centers and vice versa, from Science Data Centers to Science Data Center, from PI-managed Mission Data Centers to PImanaged Mission Data Centers, etc. are omitted for simplicity. Four different types of data flow are identified in the diagram. Internal data flow refers to data flow inside each data center. L0, or raw spacecraft instrument data flow, refers to spacecraft or level 0 data flow between mission operations. Pl-managed Data Centers or Multimission Data Centers, and Backbone or Long-Term Archive Data Centers. Distribution flow denotes data distribution to end-users. System interchange flow denotes data exchange between data centers. As suggested by Figure 1.1.1, the ESE network provides a means for opening numerous new channels for Earth Science satellite data streams to reach the user community. Such data streams will flow to users both directly from mission data processing centers as well as via many intermediate information providers.

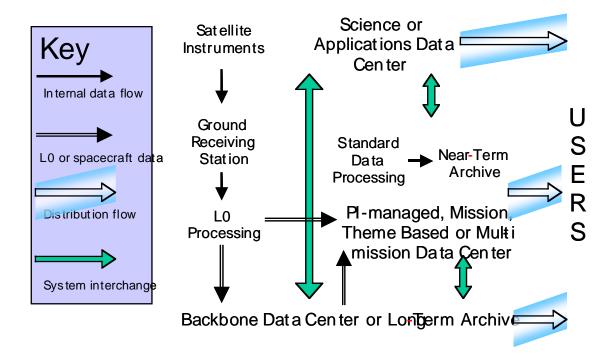


Figure 1.1.1 Simplified ESE network Data Flow (Adopted from Figure C-2 [1])

The SEEDS Near-Term Missions Standards (NTMS) study group is tasked to make recommendations for the use of standards by the ESE near-term missions. These standards are not meant to prescribe the ways that each near-term mission manages data internally or the L0, or raw spacecraft instrument data flow. Instead, the recommended standards pertain to the data distribution to end-users and to the data interchange between the ESE network of data systems and services components (*i.e.*, between different data centers as shown in Figure 1.1.1).

1.2 The Rationale for Standards

Standards aid in interoperability between data systems and facilitate access by users and the software they use. The successful adoption and use of standards for the ESE network of data systems will reduce the cost and enhance the efficiency of data system development and maintenance. Use of standards for the interchange among the ESE data and service components also makes it easy for data and service providers to join the ESE network of data systems without negotiating one-to-one agreements with each potential provider. The standards that the NTMS study group is addressing include data packaging standards, data catalog interface standards, metadata standards, and documentation standards, as defined below.

- Data Packaging Standards define how to package or encode data that is stored on a computer or transferred from one system to another. Software libraries may be available to facilitate decoding, encoding, or manipulating data packaged in a particular way.
- Data Catalog Interface Standards specify data search-and-order requests between ESE data and services components, usually over a network. These interface standards are defined independently of the data's packaging (encoding).
- Metadata / Documentation Standards provide a common lexicon and a set
 of attributes describing data to ensure that users can 1) find the data in
 catalogs, registries, and other indexes; 2) interpret the data
 unambiguously; and 3) apply system services correctly. Metadata is
 usually highly structured and formalized, whereas, documentation usually
 refers to more free-text descriptions. Most metadata and documentation
 standards are content standards (format-independent); XML is a popular
 encoding for metadata.
 - Content Standards for data or metadata define the information elements and their intended meaning (semantics), independently of how these elements may be encoded in files (their syntax). Two or more encodings of the same content standard can be mapped (machine-translated) to each other with no loss of information.

For years, various satellite missions and scientific communities have found ways to use each other's data, but stable, rich standards can further promote opportunities in research and applications for data users worldwide. The evolution of these standards over the past 25 years or so has largely been driven by specific science communities with a goal of making life easier for themselves. The past 10 years or so has seen ever wider global scientific communities tied together through the Internet with a goal of still faster-paced data exchange and hopefully faster-paced research results. However, the diversity of available data sources and data standards presents a significant challenge to Earth science researchers, especially interdisciplinary Earth scientists.

As almost any researcher can attest, a substantial portion of the resources required to perform an investigation are expended on locating, obtaining, and then reading and possibly reformatting the necessary data. Standardization of data formats, metadata, documentation, and catalog interface standards can lower the threshold on data exchange between the ESE network of data systems and services components and the user access to the data products. The Internet offers a compelling example of the essential role standards play in facilitating data exchange. Without the underpinnings of the Internet - TCP/IP, HTML, SMTP, GIF, JPEG, PDF, etc., the explosion of information exchange brought about by the Internet could never have happened.

1.3 Assumptions

This study focuses on near-term missions that are already in formulation and is aimed to provide concrete, specific recommendations for the near-term missions' use. The following assumptions are made to carry out this study.

- The emerging field of Web Services is driving rapid development of data format-neutral service interface standards. Examples relevant to ESE data include the OpenGIS Web Map Service and Web Coverage Service. However, the use of online services is still only emerging in practical ESE work; it will take some time before Web Services become a part of mainstream data access and distribution.
- For the near-term missions, the preferred mode of delivering data remains the transfer of discrete files. Therefore the file format itself is critical to the interchange standard.
- 3. Content data standards (define the information elements and their intended meaning (semantics), independently of their syntax) provide well-known semantics that can support interoperability through translators or cross-reference tables. The leading definition for such standards is the Federal Geographic Data Committee (FGDC) that has developed Content Standard for Remote Sensing Swath Data and Content Standards for Digital Orthoimagery. However in practice, content standards alone may not suffice for transferring complex data between different user communities without information loss or distortion.
- 4. The processes of standards development and adoption are the responsibility of the long-term standards study team. The NTMS study recommends a first evolutionary step in adoption of standards by endorsing specific standards and practices.

1.4 Methodology

This document provides recommendations for the use of standards by the near-term missions. We analyzed what standards are currently in use in the near-term heritage missions and other EOS missions, posing questions such as: What are the lessons learned on implementing and using those standards currently in use? What are the lessons learned from other government agencies such as NOAA? What criteria should we use to evaluate different standards? What feedback do data producers and data users have on standards? What standards do users think NASA should use in the future? Once we provide recommendations, how can the recommendations be implemented for the near-term missions? What respective activities should be supported in order to facilitate the adoption of the standards?

This report intends to answer these questions. We studied the published objectives of the assigned missions and interviewed some key planners in an attempt to understand the role data systems and data systems standards were expected to play in those missions and did play in their direct heritage. We

discussed our progress at the first SEEDS Public Workshop held February 5-8, 2002 in Greenbelt, Maryland. To verify our general understanding, we also asked for, and received, direct one-on-one feedback from the near-term missions on our draft survey. We investigated each of the standards identified by the mission heritage survey and common standards used in other government agencies and industry. We researched their technical aspects and surveyed the opinions held by potential end-users and producers. We developed a structured survey and individually interviewed many EOSDIS DAAC User Working Group members and data users and producers in NOAA. We also conducted a survey of EOS data users and producers at the 2002 NASA Science Data Processing Workshop. The gathered survey, interview, and workshop opinions were consolidated.

References:

A 6 to 10 Year Approach to Data Systems and Services for NASA's Earth Science Enterprise; Draft Version 1.0; February 2001; Section A.3.

2.0 Near-Term Mission and Heritage Mission Standards

2.1 SEEDS Near-Term Missions

The missions that SEEDS is initially targeted to support include the following eight near-term missions (Table 2.1.1).

Table 2.1.1 SEEDS Near-Term Missions

Mission Name	Phase	Anticipated Launch Date
Landsat Data Continuity Mission (LDCM)	Formulation	2006
NPOESS Preparatory Project (NPP)	Formulation	2006
Ocean Surface Topography Measurement (OSTM)	Formulation	2006
Ocean Vector Winds	Formulation	2007
Global Precipitation Measurement (GPM)	Formulation	2007
Solar Irradiance	Formulation	2007
Carbon Cycle Initiative (CCI)	Pre- Formulation	2008-2012
Total Column Ozone	Pre- Formulation	N/A

See Acronym List if needed

A summary of the near-term mission instruments, data formats, and metadata standards is described in Table 2.1.2. As shown in the table, LDCM, the first near-term mission, has already decided the data and metadata standards they plan to use for the mission data products (specified in the Request For Proposal (RFP) they released October 2001). Our recommendations for the use of data, metadata, and data interfaces in near-term missions may, or may not, impact the LDCM mission.

Table 2.1.2 SEEDS Near-Term Mission Standards

Missions	Instrument	Data Format	Metadata Format
LDCM	Not specified	1. HDF 2. GeoTIFF 3. L7 Fast Format	1. ECS 2. FGDC
	ATMS		
NPP	CrIS	N/A	N/A
	VIIRS		
OSTM (or Jason-2)	N/A	N/A	N/A
Ocean Winds	Seawinds	N/A	N/A

Missions	Instrument	Data Format	Metadata Format
	Dual Frequency Radar (DFR)		
GPM	Advanced TRMM Microwave Imager (TMI)	N/A	N/A
	Nadir-viewing Microwave Radiometer		
Solar Irradiance	N/A	N/A	N/A
	A passive spectrometer		
CCI Missions: 1. Pathfinder CO ₂ 2. Ocean Carbon	A rotating scanner telescope		
3. Low Density Biomass	A hyperspectral imager	N/A	N/A
4. High Density Biomass5. Advanced AtmosphericCO₂	A P-band SAR and an imaging laser altimeter		
	A pulsed, dual frequency, tunable laser sounder		
Total Column Ozone	Some combination of OMPS-like, TOMS-like, SAGE-like and an IR limb sounder	N/A	N/A

See Acronym List if needed

2.2 Heritage Mission Standards

Data management information for near-term missions and heritage missions is presented in Table 2.2.3.

Table 2.2.3 SEEDS Heritage Missions Data Management Information

Mission	Heritage Mission	Heritage Instrument	Production Site	Archive Site	Data Format	Metadata Format	Catalog Interface
LDCM	Landsat 1- 7	TM ETM+	EDC DAAC	EDC DAAC	1. HDF 4 2. GeoTIFF 3. L7 Fast Format	1. ECS 2. FGDC	Collection: Z39.50, Custom Inventory: EOSDIS V0 Guide Document: Z39.50
NPP	Aqua	AMSU HSB AIRS	GSFC DAAC	GSFC DAAC	HDF- EOS 4	ECS	Collection: Z39.50, Custom Inventory: EOSDIS
1/2	1/03		7			version 2	V0 Guide Document: 739 50

Mission	Heritage Mission	Heritage Instrument	Production Site	Archive Site	Data Format	Metadata Format	Catalog Interface
	Terra	MODIS	GSFC DAAC NSIDC DAAC EDC DAAC	GSFC DAAC NSIDC DAAC EDC DAAC	HDF- EOS 4	ECS	
	Jason-1	Poseidon-2 Radar Altimeter	PO DAAC	PO DAAC	Native	Custom	N/A (Data is not
OSTM	J23011-1	Jason Microwave Radiometer	AVISO	Binary	Cusiom	available to public yet)	
	Topex/Pos eidon			PO DAAC	Native Binary for low level products	Custom	Collection: Z39.50, Custom Inventory: EOSDIS
			eidon	Microwave	AVISO	AVISO	netCDF for Level 3 product
	Adeos-1	NSCAT	JPL SeaPAC	PO DAAC	HDF 4	Adapted ECS	Collection: Z39.50,
Ocean Winds	Quikscat	Seawinds	JPL SeaPAC	PO DAAC	HDF 4 BUFR	Adapted ECS	Custom Inventory: EOSDIS V0 Guide Document:
	Adeos-2	Seawinds	JPL SeaPAC	PO DAAC	HDF 4 BUFR	Adapted ECS	Z39.50
GPM	TRMM	ТМІ	GSFC DAAC	GSFC DAAC	HDF 4	ECS	Collection: Z39.50, Custom Inventory: EOSDIS
							V0 Guide Document: Z39.50

Mission	Heritage Mission	Heritage Instrument	Production Site	Archive Site	Data Format	Metadata Format	Catalog Interface
		VIIRS	GSFC DAAC	GSFC DAAC	HDF 4	ECS	
		PR	GSFC DAAC	GSFC DAAC	HDF 4	ECS	
		CERES	LaRC SIP	LaRC DAAC	HDF 4	ECS	
		LIS	GHRC SIP	GHRC	HDF 4	ECS	
	SNOE	XPS	LASP	LASP	ASCII	Custom	Collection: NSSDC Master Catalog Inventory: Custom Document: WAIS
Solar Irradiance	UARS SOLSTICE	SOLSTICE SIM	UARS CDHF and GSFC	GSFC DAAC	Native Binary Format	Native SFDU format	Collection: Z39.50,
	ACRIM III	TIM	ACRIM III SIPS	LaRC DAAC	HDF 4	ECS	Custom Inventory: EOSDIS V0 Guide
	EOS SORCE	SIM SOLSTICE XPS TIM	LASP SORCE SIP	GSFC DAAC	HDF-5	ECS	Document: Z39.50
CCI	SeaStar	SeaWiFS	GSFC DAAC	GSFC DAAC	HDF FF	ECS	Collection: Z39.50, Custom Inventory: EOSDIS
1/2	1/03		9			version 2	V0 Guide Document:

1/21/03 9

Mission	Heritage Mission	Heritage Instrument	Production Site	Archive Site	Data Format	Metadata Format	Catalog Interface
	Terra	MODIS	GSFC DAAC	GSFC DAAC	HDF- EOS 4	ECS	
	Nimbus-7	CZCS	GSFC DAAC	GSFC DAAC	HDF DSP CRTT	Native format	
	VCL	MBLA	Raytheon ITSS	EDC DAAC	Unknown	Unknown	Unknown
Total Column Ozone	Nimbus-7 Meteor-4 ADOES Earth Probe QuikTOMS	TOMS	GSFC DAAC	GSFC DAAC	HDF-4	ECS	Collection: Z39.50, Custom Inventory: EOSDIS V0, Custom Guide Document: Z39.50
	AURA	ОМІ	GSFC DAAC	GSFC DAAC	HDF-4 for Level 0 and 1 HDF- EOS 5 for Level 2 up	ECS	N/A (Data is not available yet)

See Acronym List if needed

Several observations can be made from Table 2.2.3:

1. Most of the heritage missions use the Hierarchical Data Format (HDF 4) or HDF-EOS (Earth Observing System) data formats and the EOSDIS Core System (ECS) metadata format for archiving and distribution. Note that the HDF 5 is a newer library meant to eventually replace HDF 4. HDF-EOS, used by Terra and Aqua missions, is based on HDF 4. HDF-EOS 5, to be used by the Aura mission, is based on HDF 5. Heritage missions that do not use the HDF or HDF-EOS data formats and the ECS metadata format for product distribution are the Jason-1, Topex/Poseidon, and the Upper Atmospheric Research Satellite (UARS) missions. The Jason-1 and Topex/Poseidon missions are heritage missions to the Ocean Surface

Topography Mission. UARS is a heritage mission to the Solar Irradiance mission.

- 2. Several heritage missions distribute their data products in multiple data and metadata formats. For example, Landsat missions distribute their data products in three different data formats, namely HDF, GeoTIFF, and Fast Format, and two metadata formats, ECS and FGDC. SeaWinds distributes their data products in HDF and BUFR (Binary Universal Format For Representation of data) format. The HDF format is for distributing research data products by the NASA Jet Propulsion Laboratory (JPL) Distributed Active Archive Center (DAAC), while BUFR format is used to distribute operational data products by NOAA NESDIS (National Environmental Satellite Data and Information Service).
- Data distribution formats for heritage missions consist of HDF, HDF-EOS, netCDF, GeoTIFF, Fast Format, BUFR, Binary, and ASCII. Metadata distribution formats for heritage missions include ECS, FGDC, and custom formats.
- 4. The same catalog interface standards are used across all of the heritage missions except, SNOE, a solar physics mission. Catalog interfaces used for collection metadata include the NASA Global Change Master Directory and Mercury. The GCMD supports the ISO/ANSI Z39.50 standard as well as customized search capabilities. The Mercury is a less-structured search system for Earth science data that is based in Z39.50 and internet standards such as http and ftp. The catalog interface standard used for inventory data search and order is the EOSDIS Version 0 protocol. TOMS data can also be accessed from the TOMS instrument team website using internet standard ftp. The Guide document search is based on the Z39.50 standard. Since XSP data is archived and distributed from the National Space Science Data Center (NSSDC), both the collection and inventory catalog searches are based on the NSSDC custom interfaces and standards. The XSP documentation search is based on the Wide Area Information System (WAIS) access.

3.0 Lessons Learned

This chapter presents lessons learned from past experiences with data and metadata standards used for NASA SEEDS heritage missions and NOAA missions. Some of the lessons learned pertain to past experiences with developing or implementing the standards, and others are related to past experiences with using the standards.

3.1 Lessons Learned on Implementing and Using NASA EOS Standards

3.1.1 Landsat 7

Landsat 7 data products are archived in the HDF format but distributed in three different formats: GeoTIFF, Landsat 7 Fast Format, and HDF. Based on statistics collected by the EDC DAAC [Earth Resources Observation System (EROS) Data Center (EDC) Distributed Active Archive Center (DAAC)] User Services from January 1, 2001, to September 30, 2001, most of the users ordered L-7 data either in Fast Format (46%) or in GeoTIFF (42%). Only 12% of the users ordered L-7 data in HDF format. Of the users who ordered data in HDF format, most were from international ground stations and the data product they ordered was Level 0R. HDF is the only format available for Level 0R. These statistics indicate that:

- User communities welcome multiple distribution data formats. Statistics
 have shown that users order Landsat 7 data in all three available formats
 with the majority (88%) of the users choosing GeoTIFF or Fast Format.
 This indicates that for well-developed satellite mission user communities
 such as the Landsat data user community, multiple data distribution
 formats are needed. Different users choose different data formats in their
 applications.
- Heritage mission data distribution formats play an important role. The
 reason the majority of the Landsat 7 users choose GeoTIFF or Fast
 Format may be because the Landsat 7 heritage mission Landsat 5 data
 products are distributed in Fast Format or GeoTIFF format. Thus, users
 were already familiar with those two formats. It seems natural that users
 should choose to use a format they are already familiar with rather than
 switching to a new data format, such as HDF.
- GeoTIFF data format is gaining popularity among Geographic Information System (GIS) users. Landsat Thematic Mapper (TM) data (Landsat 4-5) products have been distributed in Fast Format since 1984. EDC DAAC began distributing Landsat 5 TM data products in GeoTIFF in recent years. However, based on the statistics collected from January 1 to September 30, 2001, almost half (42%) of the users order Landsat 7 data products in GeoTIFF format. As GeoTIFF format is becoming a popular data format in the GIS user community, EDC DAAC is considering

distributing other land remote sensing data, such as ASTER (Advanced Spaceborne Thermal Emission And Reflection Radiometer) data products, in GeoTIFF format in addition to the HDF format.

3.1.2 TERRA

The flagship in NASA's Earth Observing System (EOS), Terra launched on December 18, 1999 and began collecting science data on February 24, 2000. There are five instruments onboard Terra, namely MODIS, ASTER, MISR, CERES, and MOPITT (see <u>Acronym List</u>). The data products from Terra, consisting of a great variety of ocean, atmosphere, and land data sets, are archived and distributed in HDF-EOS format as required by the EOS project. Terra metadata conforms to the ECS data model.

In the early 1990's, NASA's Earth Science Data Information Systems (ESDIS) began evaluating data format standards in preparation for the launches of the EOS satellites. In 1993, after careful consideration of over a dozen different formats, ESDIS chose the Hierarchical Data Format (HDF) for EOS standard data products. During the ECS design phase, it was realized that while HDF was a good format to use for storing data, further standardization would be advantageous. HDF provided little convention for associating spatial and temporal information with the science data itself. To enable additional standardization, the HDF-EOS data format was developed. This format adds mechanisms for storing geo-referencing and temporal information, data organization, and metadata storage.

Terra instrument teams and users have had several problems with implementing and using the HDF-EOS standard and the ECS data model.

- The HDF-EOS Grid and Swath provided a natural structure for the bulk of data derived from instruments on Terra and other EOS missions; however, there was no convention for storing individual data values. For example, in the case of one producer, real numbers are stored in 14 bits and 2 additional bits are used for a special purpose rather than using all 16 bits to store the number. The HDF-EOS library can access these data; however, translation and other application tools can have problems. If processing is to be performed on individual words or bits, errors can occur if the user is not cognizant of the storage method.
- There was no convention for packaging both HDF-EOS and HDF objects in the same file. All MODIS (Moderate-Resolution Imaging Spectroradiometer) Level 2 and 3 products are different. Even though they use HDF-EOS structures to store their primary data, many and varied vanilla HDF objects are included in MODIS standard products. MODIS also uses global and local text attributes to store non-ECS metadata rather than dumping it all into the ArchiveMetadata attributes as the HDF-EOS design calls for. This implies that software beyond the HDF-EOS library is required to access the additional attributes.

- Even though HDF-EOS provides a standard for packaging geolocation information, there was no detailed standard for actually calculating this information. For example, some ASTER products are geolocated using a geoid (geodetic coordinates) while others are geolocated using an ellipsoid (geocentric coordinates). This is not a priori obvious to data users.
- HDF-EOS has a steep learning curve. Once that hurdle is overcome, platform independence and common packaging provide convenience in access. However, scientists who are used to flat binary format complain about the complexity of HDF-EOS.
- It was a mistake to try to have one HDF-EOS profile to fit all disciplines. In Terra MODIS case, this leads to unproductive wrangling, an overly broad profile, and poor fit for some (maybe all) disciplines. The lesson learned is to develop strong discipline specific profiles and worry about crossing disciplines later.
- An important lesson learned from Terra is not to impose immature standards such as HDF-EOS. All the following are needed in no less than launch time minus three years:
 - Need an expert base before products are defined.
 - Need tools to verify proper implementation.
 - Need experienced help desk support (and more) and to help with implementation.
- There have been many mismatches between ESDT (Earth Sciences Data Type) and metadata output from MODIS production. This has led to a large number of product ingest failures at the DAACs. Quality control on the production end is lacking, and it can be traced to the poor versioning on the MODIS processing system end. There would be no problem if the MODIS processing team acquired their Metadata Configuration Files (MCFs) from installed descriptors at the DAACs. In reality, they modify the MCF locally and then send the changes to ECS. As a result, there can be mismatches between the DAACs installed ESDT and what MODIS is using. This problem has all but disappeared since the MODIS processing team is now using only the official MCFs.

3.1.3 AQUA

AQUA is a NASA Earth Science satellite mission mainly designed to study Earth's water cycle. AQUA was formerly named EOS PM, signifying its afternoon equatorial crossing time, as opposed to the morning equatorial crossing time for TERRA. Aqua will carry six instruments in a near-polar, low-Earth orbit. The six instruments are the Atmospheric Infrared Sounder (AIRS), the Advanced Microwave Sounding Unit (AMSU-A), the Humidity Sounder of Brazil (HSB), the Advanced Microwave Scanning Radiometer for EOS (AMSR-E), the Moderate-Resolution Imaging Spectroradiometer (MODIS), and the

Clouds and the Earth's Radiant Energy System (CERES). The MODIS and CERES instruments are the same as those onboard TERRA launched in 2000. The AQUA mission launched in May 2002.

The data format and metadata standards for the AQUA instrument data are the same as those for TERRA, namely the HDF-EOS and the ECS data model, respectively. Lessons learned from the AIRS instrument team (Evan Manning, AIRS principal developer) and the AMSR-E instrument team (Dawn Conway, University of Alabama in Huntsville, Lead Software Engineer for the AMSR-E Science Team) on implementing the data and metadata standards are summarized below.

- 1. In general, using the HDF-EOS standards requires a fair amount of "buy-in" and has a steep learning curve. Instrument team developers adapted, but casual users had more trouble. For example, it was relatively easy for an instrument programmer to produce the HDF-EOS files using the simple API. A lot of end-users, however, are reluctant to accept or "buy into" HDF-EOS because it is new. Both the AIRS and the AMSR-E teams found that HDF-EOS is very easy to use.
- 2. The HDF-EOS format has adequately supported AIRS and AMSR-E requirements, but:
 - The HDF-EOS should explicitly support field annotations. Without a standard, some developers will add their own annotation to internal HDF objects.
 - The field/attribute distinction is not clear. It seems that a swath attribute is anything that does not have a dimension that is a geolocation dimension. HDF-EOS Swath thinks it's anything with less than 2 dimensions.
- 3. The documentation for the HDF-EOS is clearly inadequate. It could really use some good sample programs. For example, provide examples that actually do something non-trivial, such as check for error conditions.
- 4. While AMSR-E Lead Science Computing Facility (SCF) found that implementation of the required ECS metadata was simple and straightforward; the AIRS team encountered several problems implementing the ECS data model. In fact, the AMSR-E team found the Science Data Processing (SDP) toolkit unnecessary to complete their tasks. It was noted, however, that the ECS keywords should better relate to keywords used in the GCMD (Global Change Master Directory). Problems that the AIRS team encountered are:
 - The ECS tools for implementing the ECS metadata standards are not easy to use. There are some really tricky parts, like setting "hdfattrname" to "coremetadata.0" or "coremetadata" depending on whether it is embedded metadata or not. The interface is generally confusing.
 - The amount of lead-time for adding an ECS Product Specific Attribute or changing attribute valids, etc. is too long.

- Documentation for the ECS data model is not adequate.
- The AIRS team supported ESDIS's (led by Bob Lutz) attempts to add new valids for ScienceQualityFlag. The failure of those attempts makes it hard for AIRS to support data access as they would prefer to.
- 5. On a general development note, both teams discovered the importance of regular, consistent communications (telecons, meetings, etc.) between the SCF, SIPS (Science Investigator-lead Processing System), DAAC, and ECS.

3.1.4 AURA

Aura is a NASA mission to study the Earth's ozone, air quality, and climate. This mission is designed exclusively to conduct research on the composition, chemistry, and dynamics of the Earth's upper and lower atmosphere by employing multiple instruments on a single satellite. Aura's chemistry measurements will follow-up on measurements that began with NASA's UARS and will continue the record of satellite ozone data collected from the TOMS (Total Ozone Mapping Spectrometer) missions. The satellite will be launched in January 2004 and will operate for five or more years. The Aura data products will be distributed in HDF-EOS5 format. Aura metadata will conform to the ECS data model.

The HDF file format was designed to be a very flexible format. It is able to store many different types of scientific data in a variety of ways. While this flexibility is an asset to customized data storage, it is not ideal when one is trying to ease sharing of data. As there is so much flexibility, two different developers storing the exact same data can store the data in dramatically different ways. To constrain HDF for use in the EOS community, HDF-EOS was developed.

While HDF-EOS constrains HDF with its POINT, GRID, and SWATH interfaces, it is still possible to create two files that are completely different and require dramatically different readers. Areas of potential mismatch include:

- Organization of data fields and attributes
- Dimension names
- Geolocation names and dimension ordering
- Data field names and dimension ordering
- Units for data fields
- Attribute names, values, and units

When the Aura Data System Working Group (DSWG) reviewed the proposed structure of the Level 2 data files from each instrument, it was discovered that each instrument's data files were, at times, quite different. DSWG agreed that with a little work, it was possible to adopt a uniform set of file format guidelines and that it was advantageous to do so. One of the main advantages of this standard is to allow users the ability to use the same set of tools and I/O routines for any of the Level 2 data from instruments within Aura. At the time of this

writing, the "HDF-EOS Aura File Format Guidelines" has been adopted by all of the EOS Aura instrument teams. The guidelines contain detailed, specific information on how to store data. All of the items listed above are specifically addressed. As the launch of Aura has not yet occurred at the time of this writing, the outcome of this endeavor has not been determined, but it is hopeful that by adopting a uniform set of strict guidelines that the benefits will be many. The current guidelines can be found at:

http://www.eos.ucar.edu/hirdls/HDFEOS_Aura_File_Format_Guidelines.doc (Microsoft Word version)

http://www.eos.ucar.edu/hirdls/HDFEOS Aura File Format Guidelines.pdf (Adobe Acrobat format)

3.1.5 QuikSCAT/SeaWinds

The SeaWinds instrument on the QuickScat satellite is a specialized microwave radar that measures near-surface wind speed and direction under all weather conditions and cloud cover. It was launched in 1999 as a follow-on mission to the NASA scatterometer (NSCAT) that flew on the Japanese ADEOS-1 (Advanced Earth Observing Satellite) platform during 1996-1997; and the Seasat-A scatterometer system (SASS), which flew in 1978.

A unique feature of the QuikSCAT/SeaWinds mission is that SeaWinds data are processed, archived, and distributed at both NASA JPL and NOAA NESDIS. SeaWinds data are downloaded from QuikSCAT once every orbit (101 minutes). The stream passes on from the receiving ground station to the Central Standard Autonomous File Server (C-SAFS) at Goddard Space Flight Center (GSFC). The data are then forwarded to both JPL and NOAA. JPL uses these data to produce its science-level wind product, while NOAA uses an altered version of JPL's processing to produce its own Near Real Time (NRT) wind product. This dichotomy can be summarized as follows:

- While the processing software used at NASA JPL and at NOAA NESDIS
 is the same, data products produced at JPL are research products (with
 higher accuracy) used for research and in the application community,
 while data products from NOAA are near real-time products (within 3
 hours of observation) targeted for operational users such as the National
 Weather Services (NWS).
- The SeaWinds products distributed by JPL are in HDF format while data products distributed by NOAA NESDIS are in BUFR format. This is because many operational and modeling users use the WMO (World Meteorological Organization) data standards, BUFR and GRiB (GRidded Binary). NOAA is required to provide data to their operational users in BUFR/GRiB format. For the future, the current plan is to move the NRT processing from NOAA to the Physical Oceanography (PO) DAAC at JPL, starting with the ADEOS-II mission in 2002.

3.1.6 ACRIM

For ACRIM, using HDF-EOS was required; however, since mapping the terrain of the Earth was not necessary (ACRIM is solar pointing), the EOS part did not apply. ACRIM was actually using something akin to a subset of HDF. Because ACRIM used HDF in a limited fashion, enough tools were available, but it still required the team to learn almost everything about HDF in order to determine what functions they actually needed. Overall, HDF was relatively easy to implement. Some lessons learned indicate that the following would have been helpful in the implementation of HDF:

- An instruction manual "What would have been helpful is a manual with step-by-step instructions; it could have been a quicker implementation."
- Help desk "Having someone who could spend a little time over the phone would have been very helpful."
- Rectifying the problems with creating HDF files with REAL and INTEGER values.

[Frank Boecherer, ACRIM Science Computing Facility, Personal Communication, June 2002]

3.1.7 SeaWiFS

Ten years ago, when SeaWiFS was in development, HDF had some capabilities that were not supported at that time. In the beginning, HDF was largely an image format; it only supported a limited number of data sets, and it had floating point numbers only. The SeaWiFS team identified these deficiencies early on; documented and issued reports; then received responses from National Center for Supercomputing Applications (NCSA). As a result, HDF was made more friendly and easier to use. In addition, the parallel development of HDF for use with IDL allowed users to write their own HDF tools. The main thing that was learned through the experience of implementing HDF into the SeaWiFS project was that good user support is essential. The group at NCSA responded to all of their needs at the time. "That was the thing that made it work – user support, help desk." [Fred Patt, SAIC Project Manager, Personal Communication, June 2002]

SeaDAS (SeaWiFS Data Analysis System) is a comprehensive image analysis package for the processing, display, analysis, and quality control of all SeaWiFS data products, ADEOS / OCTS (Advanced Earth Observing Satellite / Ocean Color and Temperature Scanner, Japan), MOS (Modular Optoelectronic Scanner, Germany), CZCS (Coastal Zone Color Scanner, NASA), and Ancillary data (Meteorological, Ozone). HDF facilitated the development of this powerful tool. The versatility of HDF also allows individuals to develop their own uses within the SeaDAS system. HDF was mandated for the SeaWiFS project because EOS was still under development, and SeaWiFS was to pave the way for future missions. One lesson learned is: allow time to develop tools (or

preferably use existing tools) to facilitate ease of use. [Jim Acker, DAAC User Support, Personal Communication, June 2002]

3.1.8 Jason-1

Jason-1 chose to use its own ad-hoc, format (binary) as the primary data product for historical reasons (continuity). The main advantage of using binary is that it is fast and simple. Once given the read program, it is self-contained. A disadvantage to binary is that each data set requires its own read program.

Initially, one of the problems with HDF was that software for reading the format was not widely available, and it did not work on many important computer classes. A second problem, in the past, was that installing the HDF libraries required major system administration knowledge. Also, the initial jump into HDF is difficult and requires a lot of "handholding", but only for first-time users. However, the beauty of HDF is uniformity across mission data sets.

From these ideas, the main lessons drawn are:

- Before declaring a format "STD", make sure its supporting software installs properly and runs on the main machines intended.
- Understand which classes of users will be EXCLUDED by the new format (for example, the simple binary format of Topex can be read on even a windows 95 computer, but HDF will not install there). It is acceptable to exclude classes of users CONSCIOUSLY, but not because of oversight.
- Do not underestimate the "handholding" that will be needed to help users install, then run, the new software. HDF, etc. are not 'read programs,' they compare to major operating systems or major commercial packages (IDL, Matlab, Mathematica, etc) in their complexity and their installation can be as complex.

[Victor Zlotnicki, Jet Propulsion Laboratory, Personal Communication, June 2002]

3.1.9 **AVHRR**

AVHRR data format was based on TIROS data for continuity (level 1B, native binary). However, about 2-years ago, NOAA began offering AMSU data in HDF-EOS along with the BUFR and 1B products. The response to HDF-EOS was great. Almost all of the climate scientists are now using the HDF-EOS format by their own choice. In the future, NOAA hopes to offer AVHRR as an HDF-EOS product, due to customer demand. [Ingrid Guch, National Environmental Satellite, Data and Information Services (NESDIS), Personal Communication, June 2002]

The HDF format has already been chosen for the reprocessing of all AVHRR data for JPL. It was known that the data files would need to be compressed, but the problem was, if just a small part of a big data set was needed, the entire file would have to be decompressed and then the small subset would have to be extracted. With HDF, a chunking process exists (also called tiling). This

compresses the data in such a way that it allows storage of data sets in chunks that can be decompressed separately. Thus, HDF-4 was chosen for the reprocessing of the AVHRR data. [Peter Cornillon, University of Rhode Island, Oceanography Department, Personal Communication, June 2002]

3.2 Lessons Learned on Implementing and Using other Standards

3.2.1 NOAA Standards

The National Oceanic and Atmospheric Administration's (NOAA's) National Environmental Satellite, Data, and Information Service (NESDIS) operates NOAA's environmental (weather) satellites and manages the processing and distribution of the data and images these satellites produce daily. NOAA's operational weather satellite system is composed of two types of satellites: Geostationary Operational Environmental Satellites (GOES) for "now-casting" and short-range warning and Polar-Orbiting Environmental Satellites (POES) for longer-term forecasting. Both types of satellites are necessary for providing a complete global weather monitoring system. The primary customer is NOAA's National Weather Service (NWS), which uses satellite data to create forecasts for the public, television, radio, and weather advisory services.

NOAA NESDIS does not use consistent data and metadata formats for their POES and GOES satellite data archive and distribution. The POES and GOES data are processed by the Information Processing Division (IPD) of the NESDIS Office of Satellite Data Processing and Distribution (OSDPD). The IPD is responsible for ingest, processing, and dissemination of environmental satellite data. The GOES data are distributed in McIDAS formats. The POES weather and climate data products are distributed in various different data formats including flat binary file, Level 1b, GIF, ASCII, BUFR, GRiB, HDF-EOS, netCDF, and McIDAS [1].

- In general, NOAA NESDIS uses multiple distribution data formats to satisfy different user communities' needs [Ingrid Guch, NOAA NESDIS, personal communication]. The National Weather Service or the modeling community (US and international) uses the WMO data standards, BUFR and GRiB. These users have been relying on NOAA to format the data in BUFR and GRiB (as opposed to them taking the data and running their own converter). The BUFR/GRiB formats are very complex, though, and not generally used by the people outside the modeling community.
- The imaging, climate, and scientific community as well as the NOAA NESDIS maintenance personnel greatly prefer the HDF-EOS data (ease in visualization, combining datasets, using commercial software, etc.).
 The netCDF format has the same benefit.
- Other experienced users (education, academic, etc.) seem to prefer a binary or ASCII flat file so they can easily manipulate it and add GIS or whatever extensions they like.

- Browsing users (education, some academic folks, etc.) prefer the option of ASCII, spreadsheet, and GIF.
- For satellite data (sensor counts with navigation and calibration appended but not applied), users seem satisfied with the current packed binary file (Level 1b format). The internal NESDIS maintenance personnel have been using an unpacked binary file (Level 1b star) for ease of use in realtime processing. However, this requires recreation of the "unpacked" file from archived metadata and the 1b if reprocessing is necessary (problems occurred in the real-time processing).

Long-term environmental satellite data products are archived and distributed at the NOAA National Climatic Data Center (NCDC). Archive formats used in NCDC are different for different data products. Many products are archived in a custom format and others are in HDF-EOS, Level 1b, ASCII, or JPEG [Kathy Kidwell, NOAA NCDC, personal communication, 2002]. Data distribution formats are the same as the archive formats in NCDC. Lessons learned on NOAA data standards are summarized below:

- Since NOAA is an operational agency and its main customer is the NWS, NOAA NESDIS is required to distribute their satellite data in BUFR/GRiB format to the NWS or the modeling users, although there are many problems with the BUFR/GRiB format [Ingrid Guch, NOAA NESDIS, personal communication; 2002].
- NOAA NCDC has many legacy systems and they have problems translating data to/from BUFR/GRiB format [Geoffery Goodrum, NOAA NCDC, personal communication, 2002].
- The NOAA NESDIS staff have had a positive experience with the HDF-EOS data format [2] and their users, mainly imaging, climate, and scientific communities, like the HDF-EOS format because of the flexibility, tools, and vendor support [3].

3.2.2 The Spatial Data Transfer Standard (SDTS)

The Spatial Data Transfer Standard became a Federal Information Processing Standard (FIPS 173) in 1992, after a 10-year development effort. It was to serve as the national spatial data transfer mechanism for all U. S. Federal agencies, and to be available for use by state and local government entities, the private sector, and research organizations. SDTS specifies exchange format constructs, addressing structure, and content, for spatially-referenced vector and raster data, to facilitate data transfer between dissimilar spatial database systems.[4] The Spatial Data Transfer Standard (SDTS) doesn't prescribe a single data model; rather it provides a set of rules intended to represent virtually any data model.

However, SDTS fell short of its ambitious goals; and the marketplace was slow to accept and support it. Arctur *et al.* [5] list a number of reasons for this:

Complexity - SDTS was driven primarily by large national-level data producers and their needs (very large databases, complex

interdependencies, high precision, flexible models, extensive metadata, collaborative updates, etc.). These needs far exceeded those of casual "desktop GIS" users and of most commercial, regional, or local GIS projects, and they stretch even today's GIS technology to its limits. Many people in the GIS community found SDTS to be overly complex, few understood its intended purpose, and thus few chose it when other, more established formats were available.[6] (Arctur et al. [5] suggest that as GIS users become more sophisticated, they may demand more of their technology (including data models and formats), and be more able and willing to cope with the implied complexity.)

Slow development of the standard in a fast-changing market - In the decade that elapsed between the first work on SDTS and its final adoption as a standard, the GIS industry grew significantly, and several vendor-specific exchange formats came into widespread use, which satisfied many users' immediate needs, and thus limited the community's interest in using SDTS (which many perceived as yet another format). Even though the standard was mandated for all federal agencies, most data suppliers, responding to user demand, offered alternative data encodings – and only the most curious and experimental users chose SDTS.

Limited vendor support - SDTS got caught in a "chicken-and-egg" situation with GIS vendors: in order to build market demand for SDTS-aware software, data providers needed to produce large volumes of SDTS data. But they needed to use commercial GIS products to build these data; so they had to persuade vendors to produce SDTS products in the absence of customer demand. A few vendors did include STDS conversion tools in their products (e.g., ESRI's Arc/Info, Laser-Scan's Gothic); however different products interpreted SDTS ambiguities differently (see below), so they would often fail to translate unexpected STDS constructs introduced by another vendor's product.

Slow development of practical profiles - SDTS was a very general standard: any practical use of it required users to agree on a particular profile. But due to the complexity of SDTS, and the limited educational material (such as usage examples) available to the geospatial community, it took another four years to complete the first usable profile of SDTS (the Topological Vector Profile). The lack of interest in, and understanding of, SDTS among the GIS community also reduced the demand for useful profiles, and the community's enthusiasm for working on them. In the end, this first profile proved to be both limiting (encoding fairly mundane examples required awkward workarounds) and unnecessarily complex (it required arc/node/polygon topology, which was unnecessary or even meaningless for many commonly-used cases). [7]

Harmonization delays - Subsequent efforts to define other SDTS profiles (the Raster Profile and Transportation Network Profile) were almost complete when they became mired in attempts to harmonize them with similar standards being developed in NIMA, NATO, and the European

Union. This resulted in further delays to their development. (Arctur *et al.* [5] suggest that early harmonization is easier, and that profiles should not be developed so quickly as to overlook other, related standards.)

Ambiguity in the data model (e.g., the cardinality of relationships) and the data semantics (e.g., the meaning of relationships among entities) of SDTS and its profiles limited the utility of SDTS for reliable information transfer. (Arctur [8] likens an SDTS profile to a game in which teams agree on the size of the ball and the shape of the field, but not on the rules of play.) SDTS was supposed to be very general, and to make datasets self-describing; that is, the data model could be determined from the dataset contents. But this proved an elusive goal; and thus many even of those who were willing to be "SDTS pioneers" ultimately concluded that its practical value was limited.

In addition, during and after the development of SDTS, new, unanticipated technical expectations arose, which demanded significant technical (re)design and international coordination, and further weakened the community's support for SDTS:

- a standard means of representing subtiles within a dataset;
- support for permanent, universally unique object identifiers across all datasets:
- support for value-added extensions and incremental updates by users;
- support for tracking changes and historical lineage of features and spatial primitives;
- harmonizing the metadata content with emerging international standards;
- harmonizing repository organization with emerging OpenGIS software interfaces.

Some of these issues might have been anticipated in the design of SDTS, while others stemmed from the increasing sophistication of GIS products and their users over the years.

The need for harmonization with OpenGIS led to OpenGIS' work on interface specifications for access to geospatial data (features, coverages, identifiers, etc.). Since the late 1990s, OpenGIS has been the locus of much subsequent work in this area. It focused first on accessing geospatial data (e.g., Simple Features Access for SQL, COM, and CORBA), then on encoding geospatial features in XML (Geography Markup Language (GML)) for transfer between clients and servers.

In summary, the SDTS experience illustrates the importance of keeping pace with technology and market trends and emerging expectations, even after capturing initial requirements. It shows the role of timing: a standard may be "ahead of its time" (arriving before people are ready to understand them or accept more complexity) or "overcome by events" (arriving after people are used

to making do without flexible, general, or vendor-independent solutions). Paradoxically perhaps, SDTS was both!

The SDTS experience also underscores the need to balance advanced needs with more basic ones; the importance of good documentation and usage examples; the challenge of "priming the pump" among vendors in advance of market demand; the benefits and risks of harmonizing with related standards; and the futility of mandating a standard that fails to meet a need.

References:

- [1] NESDIS Satellite Product Overview Display, http://osdacces.nesdis.noaa.gov:8081/satprod/products/prod_frameset.cfm?prodid=-1
- [2] Huan Meng, Doug Moor, Limin Zhao, Ralph Ferraro, HDF-EOS at NOAA/NESDIS; Presentation; HDF-EOS Workshop, 2000, Landover, MD: http://hdfeos.gsfc.nasa.gov/hdfeos/WSfour/meng/hdfeos4.ppt
- [3] Andrew S. Jones and Thomas H. Vonder Haar; A Dynamic Parallel Data-Computing Environment for Cross-sensor Satellite Data Merger and Scientific Analysis; Accepted for publication by Journal of Atmospheric and Oceanic Technology; March 2002.
- [4] Fegeas, Robin (1995). Q3.3: What is this SDTS thing and is it available via ftp? In GIS-L / comp.infosystems.gis Frequently Asked Questions and General Info List (Lisa Nyman, ed.). http://www.prenhall.com/startgis/fag.html.
- [5] Arctur, D., Hair, D., Timson, G., Martin, E., and Fegeas, R. (1998) Issues and Prospects for the Next Generation of the Spatial Data Transfer Standard (SDTS). *International Journal of Geographical Information Science* 12(4): 403-425.
- [6] Hastings et al. (1996) *The Spatial Data Transfer Standard: Closing the Loop?* Panel Discussion at GIS/LIS--Denver, Colorado--November 19, 1996. http://www.ngdc.noaa.gov/seg/tools/sdts/gislis_main.html
- [7] Kelley, C., and Gosinski, T., 1994: Spatial Data Transfer Standard: do you fit the profile? *GIS World*, August 1994, pp. 48-50. http://wwwsgi.ursus.maine.edu/gisweb/spatdb/urisa/ur94076.html
- [8] Arctur, David K., 1996: Spatial Data Transfer Standard: A GIS Vendor's Perspective. Online article. http://web.archive.org/web/19990222153931/www.lsl.co.uk/~arctur/portfolio/sdts.html

4.0 Essential Standards Concepts

Before evaluating individual data or metadata standards, it may be useful to review several key concepts crucial to understanding and comparing standards.

- A comparison with private, ad-hoc, binary information transfer
- Mandatory vs. optional elements of a standard; profiles and extensions
- Abstract vs. implementation standards
- Content and format standards vs. behavior and interface standards

4.1 A Comparison with Private, Ad-hoc, Binary Information Transfer

Webster's dictionary defines a standard as: "That which is established as a rule or model by authority, custom, or general consent." Thus, standards exist only within a community of people sharing certain usage patterns ("custom") or organizational structures (formal "authority" or informal "general consent"). The emphasis in this study is on standards accepted by a fairly broad set of users, publicly documented, and either stable (unchanging) or changeable only by a consensus among these users.

Another aspect of standards is that they govern only a part of the information transfer process. For instance, GeoTIFF codifies the georeferencing of an image, but is silent on the meaning of its pixel values. Whatever a standard does not specify is left to the private (often implicit) understanding of each user community or to ad-hoc ancillary information (such as a README file or a telephone message describing data details). So, at one extreme, complex and rigid standards specify every aspect of information transfer, and at the other extreme, private agreements or ad-hoc communications leave everything implicit or unstructured. Most standards fall somewhere in between; they govern a certain piece of the information transfer process to let a certain set of users communicate or work together, but users must also rely on other standards, private agreements, or ad-hoc qualifiers.

Many earth science data users favor a "raw binary" data format that is both simple and comprehensive. In fact, "raw binary" doesn't actually mean mysterious data files that one must guess at — but, rather, a simple format (often some kind of raster grid) used by a small set of colleagues with little attention to documentation or stability. Thus, "raw binary" denotes not a single format, but as many different formats as there are workgroups. Each such format has a syntax and semantics invented just for that data set or data series, usually without a lot of attention to other related formats, and with many details left implicit or provided separately in a mission report or some other natural-language document.

Such a format may serve many people's immediate needs, for several reasons.

• A given science team often works with only one kind of data and so gets used to the one syntax for that data (e.g., keeps reusing the same parser)

- and the one set of semantics (*e.g.*, pins a page from the mission report to the cubicle wall).
- Science teams traditionally put a low emphasis on making their data accessible to others outside of their immediate colleagues. They may feel that they have done their job by distributing a simple "README" file with the data.
- ESE data is commonly encoded as raster grids for which one can make easy guesses as to syntax (band-sequential, etc.).
- Most importantly, perhaps, the semantics of much ESE data (platform orientation, sensor model, calibration information, interpretation algorithms) are so complex that bundling them with the data is often difficult; so they tend to remain in a mission report document or in people's heads. (For example, each MODIS L1b granule has dozens of ancillary data items required for proper interpretation along with several grids of data error and reliability estimates.)

However, the use of raw binary data relies much more on private agreements among colleagues than on documented, consensus standards. It has many of the properties opposite to those of standards, as listed above in the "Rationale for Standards" paragraph. It limits the ability of science teams to move beyond traditional work methods towards more effective interdisciplinary research, collaborative work, and applications. The essential points are:

- Data in a standard formats (should) convey something about the data that its users need to know; whereas, users of binary data must rely on inside knowledge or educated guesses to read and interpret the data.
- Data in a standard format may be used outside of an "inner circle" of colleagues, but only holders of the necessary private information can use raw binary data.
- Standard data formats limit the need for pair-wise translators to and from every possible format, whereas, each raw binary format needs a different translator.
- Standard data formats, by fixing the syntax and semantics of information, allow the possibility of machine-to-machine communication between different systems (that is, interoperability). In contrast, raw binary data requires human inspection an intervention, thus, hindering (preventing) system interoperability.
- Standard data formats facilitate unambiguous transfer of information between users of different systems working with different datasets. This is more difficult with raw binary data, which often loses all but raw pixel values in translation.

In summary, the use of private agreements does not constitute a standard, and so "raw binary" data formats cannot be compared alongside open consensus standards.

Of course, members of a community may choose to turn a private, internal convention into a standard for a wider community by documenting and publishing their shared syntax and semantics and by sticking to what they document (that is, submitting any changes to a public consensus process or formal authority). Most standards are born this way when a usage community publishes its internal conventions to facilitate collaboration with others.

4.2 Mandatory vs. Optional Elements, Profiles and Extensions

Many standards have a set of mandatory elements to ensure basic interoperability plus a set of optional elements to serve a diversity of users and uses. This provides a "base" standard from which a particular community of users may define a *profile* (a more specific standard) to support richer communication among themselves, or more fine-grained control of each other's services. A profile is a standard derived from a base standard by adding restrictions: it may require (or exclude) an element that is optional in the base standard; it may limit the valid entries under a heading; it may fix the cardinality of a repeating element; and so on. But, the profile cannot contradict the base standard; anything mandatory in the base standard remains mandatory in the profile. Thus, any product that complies with the profile will comply with the base standard. [1, 2]

One profile presented here is HDF-EOS, an EOS-specific adaptation of the very general Hierarchical Data Format. Of the many different file structures that are possible with HDF, HDF-EOS defines three (Point, Grid, and Swath), each with spatial and temporal details alongside scientific data. In another example, FGDC's Metadata Content Standard has allowed several community-specific profiles to be defined, and in fact, the International Organization for Standardization's (ISO) Metadata standard was designed primarily for profiles. It defines several hundred elements of which fewer than 20 are required; the remaining elements are shared vocabulary (*i.e.*, a dictionary) for building profiles.

Related to profiles is the notion of *extensions*. These are elements added to the base standard by consensus among a certain community of users. As with profiles, extensions do not contradict the base standard – what's mandatory remains mandatory; products that fit the extended standard have everything needed by the base standard, and more. Nonetheless, by adding more loosely controlled, loosely defined elements to a standard, extensions may complicate the interoperability and maintenance of the standard.

For example, the earth imagery user community has extended the FGDC metadata content standard to more fully describe remotely sensed data by adding metadata elements such as the sensor model and the orbital platform, both of which the base standard doesn't provide [4].

4.3 Abstract vs. Implementation Standards

Standards and specifications for information systems are defined primarily at two different "levels of abstraction;" implementation specifications and abstract models [5].

- Implementation specifications tell software developers how to express
 information or requests within particular distributed computing environments
 (such as XML, Java, or the World Wide Web). Such standards define data
 formats, access protocols, object models, naming conventions, etc., in terms
 that are directly usable within the targeted computing environment.
 - Implementation specifications are the more immediately useful standards when they apply to one's chosen computing context. The data-format standards are implementation specifications, as are the eXtensible Markup Language (XML) encodings of FGDC, ISO, and other metadata standards.
- Abstract models specify what information or requests are valid or required in principle, irrespective of individual computing environments. They define the essential concepts, vocabulary, and generic structure (type hierarchy) of computational services and information transfer. Although not directly usable to build data or software, these models set the stage for creating implementation specifications and for extending existing ones to new environments.
 - Abstract models provide well-known semantics that can support interoperability through translators or cross-reference tables. For instance, thanks to FGDC's Content standard, Z39.50's GEO profile can "normalize" any FGDC compliant metadata (regardless of actual record formats or field names) for external access – that is, map its internal data elements to the GEO field names for external access.
 - o In general, consensus-based abstract models of data are often termed "content standards." They define the information elements and their intended meaning (semantics) independently of their syntax – that is, independent of how these elements may be encoded in files on disk or along a communications link. In principle, content standards allow different parties to communicate meaningfully by mapping their data element names to those of the content standard even when they use different formats for their data. This works well for fairly simple data structures such as the "parameter=value" pairs of many metadata files and catalog records. However, with more complex syntax or semantics, translating the abstract concepts of the content standard into the terms of a particular format often becomes an interpretation task requiring judgment calls, assumptions, and ambiguity. So in practice, content standards alone may not suffice for transferring complex data between different user communities without information loss or distortion.

4.4 Content and Format vs. Behavior and Interface

Table 4.4.1 shows that at each level of abstraction certain standards define the *interfaces* that allow different systems to work together or the expected *behavior* of software systems. This is the computation viewpoint, whose accent is on invoking services effectively and unambiguously. Other standards define the *content* of geospatial information or its *encoding* (or packaging) for accurate transfer between different processing systems. This is the information viewpoint, which emphasizes efficient, lossless communication [5].

Table 4.4.1 Viewpoints and Levels of Abstraction

	Service Invocation (computation viewpoint)	Information Transfer (information viewpoint)
Implementation specifications ("how")	Interface	Encoding (format)
Abstract models ("what")	Behavior	Content

For distributed computing, both of these viewpoints are crucial and intertwined. For instance, information content isn't useful without services to transmit and use it. Conversely, invoking a service effectively requires that its underlying information be available and its meaning clear. However, the two viewpoints are also separable: we may agree on how to represent information regardless of what services carry it; conversely, we may define how to invoke a service independently of how we package the information needed or conveyed by the service.

In a given context, either the computation view (implemented as interfaces) or the information view (implemented as formats) may take precedence. Tables 4.4.2 and 4.4.3 below show a few guidelines for prioritizing standards definition or adoption in certain contexts. In general, however, deciding which view to emphasize in a given setting is not straightforward.

Table 4.4.2 Criteria For Format Standards

Worry about a data format standard when	Don't worry about a data format standard when
Users of different formats need to share or communicate data with each other.	There's no reason for users of different formats ever to share information.
Each user group (or each user) uses a different format.	A user consensus already exists on one or a few non-proprietary data formats.
Available formats fail to convey all the information needed for proper use. (Thus users have to rely on implicit knowledge or ad-hoc notes to use the data.)	A practical, reasonably simple data format conveys all of the information users need.

Table 4.4.3 Criteria For Interface Standards

Worry about a service interface standard when	Don't worry about a service interface standard (i.e., rely on FTP/FedEx) when
Most users want the output of a few well-known processing operations, such as subsetting, filtering, transformations, etc.	Most users need direct access to raw data (as archived) for ad-hoc processing and analysis.
The intended applications are streamed or interactive – they only use parts of the available data at a given moment.	Most use of the data requires all of it (full size and detail) to be present simultaneously.
No one reasonably simple format will ever meet everyone's needs. (A service allows users to request the data they need in a format that fits it.)	Users have not begun to map their workflow to online database transactions or Web services.

Among the data standards reviewed in this report, GeoTIFF, Landsat Fast Format, and BUFR/GRiB are clearly file format standards; they specify an encoding and are silent on what access interface to use. HDF, HDF-EOS, and netCDF provide a software library to facilitate reading and writing data files, but they too are file format standards; they don't specify a format-neutral interface to a service. Table 4.4.4 compares the data models and software access libraries for a variety of data packaging standards.

Table 4.4.4 Data Models and Software Access Libraries

Data Format	Logical Model	Physical Model	Software Access Libraries
HDF	 Disk format, hierarchical, and similar to Unix file systems Self-description provided in global and local (individual objects) attributesHeader describes disk structure with metadata & pointers Usable for general scientific data storage; HFD4 data model contains: arrays, tables, raster image and text objects. HDF5 data model has HDF4-type objects imbedded within arrays and text attribute objects. Will support extended (multiple machine) files 	 XDR-based Storage layout is contiguous (serial) or chunked (direct access) Datasets consist of header attributes & data Machine-independent 	• C, C++, FORTRAN, Java
HDF- EOS	 HDF-based: Versions 4 and 5 Provides standard for geolocation data map to science data . 	 (Same as HDF) XDR-based Storage layout is contiguous (serial) or chunked (direct access) 	• C, C++, FORTRAN, Java

Data Format	Logical Model	Physical Model	Software Access Libraries
	 Point Structure: model for sparce, randomly geolocated data Swath Structure: model for data best organized by time, latitude or track parameter Grid Structure: model for data organized spatially and projected. 	 Datasets consist of header attributes & data Machine-independent Disk format is available to user 	
netCDF	Self-describingUsable for general scientific data storage	 XDR-based Storage layout direct access indexed Datasets consist of header & data Machine-independent Disk format is hidden 	C, FORTRAN, Java, Perl, Python, Ruby. Tcl/Tk
GeoTIFF	 TIFF-based, with geolocation tags Raster image data only Multiple images can be stored in a single file. Version 2 will support extended files 	 Storage layout allows random access to pixels by band, strip, or tile 	C, Perl, Python, Java
BUFR	 Tailored to atmospheric data – point data Based on sequential, tape format 	Storage layout is serialDataset consists of header + data	• FORTRAN 77
GRiB	 Tailored to atmospheric data – gridded data Based on sequential, tape format 	 Storage layout appears to be serial – "messages" Dataset consists of header + data 	Command-line translators to ASCII or IEEE binary
Fast Format	Multi-band image data	Separate header and data filesDirect access to individual bands	 Users write their own software based on examples
Binary	 Data model chosen by user. Record, data types determined by specific platform. 	Different for every productMachine dependent	Custom softwareUsers must write their own

See Acronym List if needed

4.5 Web-based Data Service Standards

The World Wide Web is driving rapid development of format-neutral service interface standards. Examples particularly relevant to ESE data include the

OpenGIS Web Coverage Service [6] and Web Map Service [7] and the Distributed Oceanographic Data System (DODS) [8].

The OpenGIS Consortium (OGC) Web Coverage Service (WCS) is likely to become an OGC specification in early 2003. It will provide access to images, imagery collections, and other systematic "fields" of values or measurements – usually arrayed on a 2D or 3D spatial grid. It fully describes the data's spatial location and its semantic content and allows clients to request subsets in space or along any of the data dimensions using a syntax based on either Uniform Resource Locators (URLs) or structured XML messages. The EOSDIS Core System (ECS) Synergy effort intends to provide WCS access to its large online data holdings ("data pools"); and the GLOBE educational project ("global learning and observations to benefit the environment:") has begun experimenting with WCS and WMS (next).

The OGC Web Map Service (WMS) provides access to rendered maps and pictures using a simple, spatial query syntax and common graphics formats (PNG, JPEG, etc.). Since its inception in early 2000, this interface has seen widespread implementation by many vendors, laboratories, and open-source efforts.

The Distributed Oceanographic Data Service provides format-neutral access to scientific datasets; its query syntax allows for "slicing" or "sampling" a dataset along any of its variable values. DODS originated at MIT and the University of Rhode Island (URI) in the mid-1990s; since then, it has seen a fair bit of implementation in the oceanographic community and among NASA DAACs. Recently, URI and NASA-DAACs have built "gateways" from DODS to WMS and WCS; and URI has begun defining two distinct successors to DODS: an "Open Source Project for a Network Data Access Protocol" (OPeNDAP) (tools for generic infrastructure protocols) and a "National Virtual Ocean Data System (NVODS)" (to supply oceanographic data and applications). [9]

(Notable Web-based services in the ESE environment include the University of Maryland's MOCHA project ("Middleware based on a code-shipping architecture") [10]; the Tropical Rainforest Information Center (TRFIC) at Michigan State University [11]; EOS-Webster at the University of New Hampshire [12]; and many others. However, these are not service interface standards but, rather, particular implementations of distributed systems. Although they provide a useful benefit to their users, they are not linked by a well-defined, published service interface standard; instead, they rely on tightly coupled components or on unpublished or proprietary interfaces.)

Finally, a number of vendors in the world of e-commerce have championed the notion of "Web Services" [13] consisting of the Web Services Description Language (WSDL) [14]; Simple Object Access Protocol (SOAP) [15]; and Universal Description, Discovery, and Integration (UDDI) [16]. These industry specifications have gained broad visibility and offer a lot of promise for Webbased data access; however, the dust is far from settling on this very active area of technology development.

In generally, the use of Web-based services is still only emerging in practical ESE work. The primary mechanism for information interchange in the ESE context remains the transfer of discrete files; it will take some time before Web-based services become a part of mainstream data access and distribution in ESE. Accordingly, this document treats format and content standards only for the near-term missions.

References:

- [1] ISO TC211 (2001), "Geographic Information Profiles" http://www.isotc211.org/protdoc/211n1134/211n1134.pdf
- [2] Federal Geographic Data Committee (1998), Content Standard for Digital Geospatial Metadata: http://www.fgdc.gov/metadata/csdgm/ : Appendix D; Appendix E.
- [3] Simon Cox (2001), Summary of some geospatial metadata standards: http://www.ned.dem.csiro.au/research/visualisation/metadata/geospatial/
- [4] NASA (2001), Digital Earth Reference Model v0.5: http://www.digitalearth.gov/derm/v05/
- [5] FGDC (2001), Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata: http://www.fgdc.gov/standards/status/csdgm_rs_ex.html
- [6] Evans, John D. (2001), OGC Web Coverage Server (WCS), Discussion Paper #01-018: http://www.opengis.org/techno/discussions/01-018.pdf
- [7] de La Beaujardière, Jeff (2001), OGC Web Map Server Interface Implementation Specification, version 1.1.1: http://www.opengis.org/techno/specs/01-068r3.pdf
- [8] Distributed Oceanographic Data System (DODS): http://www.unidata.ucar.edu/packages/dods/
- [9] Cornillon, Peter (2002), "DODS: OPeNDAP providing plug-and-play interoperability in a distributed data system," presentation at the 9th Assembly Meeting of the ESIP Federation, University of Maryland, College Park, May 15-17, 2002.

 $\frac{\text{http://www.esipfed.org/business/library/meetings/9th fed meeting/ppt/DODS.PP}}{\underline{T}}$

- [10] The MOCHA project: Self-Extensible Middleware Architecture. http://www.cs.umd.edu/projects/mocha/
- [11] <u>Tropical Rain Forest Information Center (TRFIC)</u>. http://www.bsrsi.msu.edu/trfic/
- [12] EOS-WEBSTER: Earth Science Data from the University of New Hampshire. http://eos-webster.sr.unh.edu/

- [13] World Wide Web Consortium (W3C), 2002: Web Services home page. http://www.w3.org/2002/ws/
- [14] World Wide Web Consortium (W3C), 2002: Web Services Description Language (WSDL) Version 1.2: W3C Working Draft 9 July 2002. http://www.w3.org/TR/2002/WD-wsdl12-20020709/
- [15] World Wide Web Consortium (W3C), 2002: SOAP Version 1.2 Part 1: Messaging Framework. http://www.w3.org/TR/soap12-part1/.
- [16] Universal Description, Discovery, and Integration (UDDI) of Business for the Web. http://www.uddi.org

5.0 Standards Evaluation

In order to objectively assess the data and metadata standards identified in Chapter 2 for the SEEDS near-term missions, an analysis is carried out to evaluate the standards according to many features or criteria. Furthermore, a user opinion interview/survey is conducted to gather user community's feedback on using the standards.

5.1 Evaluation Criteria

Many features or criteria can be used to evaluate the data and metadata standards identified. The intention of this study is not to identify one all-purpose standard but, rather, to identify appropriate use of the standards. For example, some standards are more suitable for transmission and archiving while others for analysis. For transmission and archiving, the most important features standards should have are semantic completeness, portability, self-description, extensibility, interoperability, etc [1]. For analysis, standards should have features such as ease of use, analysis tools support, etc. Many of these features and others are defined below.

1. Interoperability – Tools exist to translate to other standard formats with no information loss.

Is there a defined relationship or semantic equivalence between this standard and other standards? *i.e.*, can the standard be broken into elements that have the same content as elements for other standards?

Is the definition sufficiently precise to allow development of a translation algorithm between standards?

What translation tools (well known) have been developed?

2. Availability – Source code for writing and reading data in the format is widely and publicly available.

Is the source code for writing and reading data widely and publicly available? Is the software for reading and writing well documented?

Are the search and order methods for data using the format well understood and established?

3. Portability – Data in this standard can be used on a variety of platforms or in a variety of applications (vendor support).

Is the format sufficiently well defined so that data can be ported to new commonly used platforms with minimal effort?

Is the format sufficiently well implemented that new applications can access the implementation with minimal effort?

Can the standard be implemented on one platform and installed and tested on other platforms with minimal modification of source code? *i.e.*, machine dependent code is minimized.

4. Evolvability – A clear process for maintaining and evolving the standard exists.

Is there a methodology for adding new features to the standard?

Is there a software development process?

Is there a standard for documentation?

Is there an open process for evolution?

5. Extensibility – Support for extensions and profiles exists.

Does the standard allow extensions or profiles to be developed?

Are there extensions or profiles developed for the standard?

6. Self-describing - Files contain data descriptions along with the data.

Can data in this format be read without a separate document detailing file contents?

Can the data be described internally to facilitate development of applications?

Does the format contain information to allow geospatial, temporal, and/or spectral subsetting?

7. Tools Support – Software tools are available to support the standard.

Does the standard have freeware support?

Does the standard have COTS (Commercial Off-The-Shelf) software support?

8. Completeness – The capacity to carry semantic descriptive elements of the data explicitly and unambiguously. Higher levels of completeness can reduce the user's dependency on outside information, implicit knowledge, or guesswork when interpreting and applying the data.

Can the format carry everything users need to use the data correctly? *i.e.*, can the format convey the data's precise spatial location, its units of measure, the observation parameters (*e.g.*, spectral bands), accuracy estimates (error bars), and other elements needed to understand the data and apply it?

5.2 Data Standards Evaluation

Using the standards evaluation criteria defined above, Tables 5.2.1 through 5.2.8 analyze and compare data standards in use in heritage missions and other ESE missions.

Table 5.2.1 Data Standards Interoperability

	Evaluation Questions		
Data Standard	Is there a defined relationship or semantic equivalence between the standard and other standards?	Can translation algorithms be developed easily?	What Translation Tools (well known) developed?
HDF	Yes. Since HDF can contain general scientific data, it encompasses all the other standards.	Yes, HDF has a well-documented software API.	GIF <-> HDF5 HDF4 <-> HDF5 Ensight6 -> HDF5
HDF-EOS	Yes. As a superset of HDF, it also encompasses the other standards.	Yes, Point, Grid Swath add-on structures are well-documented.	GIF <-> HDF-EOS5 HDF4 <-> HDF-EOS5 Ensight6 -> HDF-EOS5 HDF-EOS -> GeoTIFF HDF-EOS Swath -> HDF- EOS Grid
GeoTIFF	Yes, for image-based standards; no, for non-image standards.	Yes. Public domain API library partially documented.	Lots of converters for TIFF; also GeoTIFF tag read & write Specialized converters for
	Na	No No ADI or libror.	L7, MODIS, MISR, ASTER
Fast Format	No	No. No API or library exists.	No
Native Binary	Depends on the standard. Most are specific to the application.	Depends on the standard, but usually not, unless specific efforts are made to document and publish an API.	No, You have to write your own translation tool
netCDF	Yes. Since netCDF can contain general scientific data, it encompasses all the other standards.	Yes. Net CDF has a well-documented API.	-> HDF -> Matlab5
BUFR/GRiB	Yes – translation of meteorological parameters to other formats is possible, with no loss of content. No for non-meteorological standards.	Yes	BUFR -> CDF

Table 5.2.2 Data Standards Availability

		Evaluation Questions	
Data Standard	Source code for writing and reading data widely available?	Read/write software well documented?	Format well described to facilitate application development?
HDF	Yes	Yes	C, C++, Fortran, and Java interfaces exist. Applications must use one of these interfaces to access the data
HDF-EOS	Yes	Yes	C, C++, Fortran, and Java interfaces exist. Applications must use one of these interfaces to access the data
GeoTIFF	Open source libraries; many COTS and freeware applications available	User interface well documented	TIFF format well documented. COTS venders sometimes use variations of the standard.
Fast Format	No	No	No
Native Binary	Not always	Not always	Not always
netCDF	Yes (C, C++, FORTRAN, Perl)	Yes	Yes
BUFR/GRiB	There are few slightly different read and write software from different organizations or countries	Not always	Not always

Table 5.2.3 Data Standards Portability

	Evaluation Questions		
Data Standard	Portable among commonly used platforms?	Format is sufficiently well implemented that new applications can access the implementation with minimal effort?	Standard can be implemented on one platform and installed and tested on other platforms with minimal modification of source code?
HDF	Precompiled HDF libraries for a variety of popular platforms such as AIX, Cray HP,SGI,Sun, Linux and Windows.	Yes	Yes

		Evaluation Questions	
Data Standard	Portable among commonly used platforms?	Format is sufficiently well implemented that new applications can access the implementation with minimal effort?	Standard can be implemented on one platform and installed and tested on other platforms with minimal modification of source code?
HDF-EOS	Precompiled HDF-EOS libraries for a variety of popular platforms such as AIX, HP, SGI, Sun, Windows, and Linux.	Yes	Yes
GeoTIFF	Works on common OSs (Linux, Unix, Windows). Designed to be portable, but need some knowledge of specs.	Need some knowledge of the specs. Need understanding of geotags to develop applications.	Yes
Fast Format	Yes	No	Yes
Native Binary	Usually not	No	No
netCDF	All major OS's: Winx, Unix, Linux, MacOS	Yes	Yes
BUFR/GRiB	YES	A generalized application would require in depth knowledge of all variants, which is not easy to obtain	YES

Table 5.2.4 Data Standards Evolvability

		Evaluatio	n Questions	
Data Standard	Is there a methodology for adding new features?	Is there a software development process?	Is there a standard for documentation?	Is there an open process for evolution?
HDF	NCSA is a currently active and outside funded group whose purpose is devoted to the HDF project. They manage development schedules and are open to suggestions from users. They are funded from a variety of sources.	Yes, HDF library is funded and developing software.	Yes, HDF library follows an internally defined standard for their documentation.	Yes, HDF group allow input from outside users

		Evaluatio	n Questions	
Data Standard	Is there a methodology for adding new features?	Is there a software development process?	Is there a standard for documentation?	Is there an open process for evolution?
HDF-EOS	Support is a contract from NASA. They respond to suggestions from users. It is NASA's decision on how long to support the contract and whether to supply money for development as well as maintenance.	Yes, HDF-EOS library is funded and developing software.	Yes, HDF-EOS library follows an internally defined standard for their documentation.	Yes, HDF-EOS group allow input from outside users
GeoTIFF	Maintained by JPL. No formal process, <i>i.e.</i> , Standards committee. The standard can be modified by others.	Yes	Yes	OpenGIS, but no formal process. Work on the GeoTIFF v2.0 spec has been slow recently, with some recent efforts
Fast Format	No	No	No	No
Native Binary	No	No	No	No
netCDF	Yes, through Unidata	Yes	Yes	Yes, informally through Unidata
BUFR/GRiB	YES	NO	Appears so, WMO issues Tech. Docs. on these formats	The WMO CBS approves changes to the format and maintains a software registry
		on Agranum List if	noodod	

See Acronym List if needed

Table 5.2.5 Data Standards Extensibility

	Evalu	ation Questions
Data Standard	Does the standard allow extensions or profiles to be developed?	Are there extensions or profiles developed for the standard?
HDF	Yes	HDF-EOS is a profile which was developed.
HDF-EOS	This is a profile of HDF	No
GeoTIFF	Yes. New projections can be added. Multiple-band GeoTIFFs allowed. GeoTIFF 2.0 will allow external files.	None that are not part of unofficial list of projections
Fast Format	No	No
Native Binary	No	No
netCDF	Yes	Yes, e.g., MINC: (Medical Image netCDF)
BUFR/GRiB	Yes	Not sure

Table 5.2.6 Data Standards Self-Describing

		Evaluation Questions	
Data Standard	Is data able to be stored so that it can be read without a separate document detailing file contents?.	Can the data be described internally to facilitate development of applications?	Does the format contain information to allow subsetting?
HDF	Data can be stored so that it is self-describing. There are no restrictions in the standard though to prevent developers from using names such as Variable1.	Data can be described with enough detail to allow applications to process data appropriately. For instance, scale factors may be included but it is developer dependent on how to do this. As a result, generic applications are limited in their scope. Applications developed for a specific data set can be very precise.	Yes, information can be supplied to allow subsetting, but there is not a requirement to do so in a consistent way. Subsetting by selecting selected data fields can easily be done on any HDF file.

	Evaluation Questions		
Data Standard	Is data able to be stored so that it can be read without a separate document detailing file contents?.	Can the data be described internally to facilitate development of applications?	Does the format contain information to allow subsetting?
HDF-EOS	Data can be stored so that it is self-describing. There are no restrictions in the standard though to prevent developers from using names such as Variable1.	Data can be described with enough detail to allow applications to process data appropriately. For instance, scale factors may be included but it is developer dependent on how to do this. As a result, generic applications are limited in their scope. Applications developed for a specific data set can be very precise.	Because of the profile, subsetting along certain geolocation fields can be done. Individual developers can break this process by not following the profile (there is no internal checking done).
GeoTIFF	Geotags and image specs.are in an ASCII header. Need library to access contents.	Information can be extracted at a pixel level. Geolocation and image info. is available through the interface.	Yes
Fast Format	No	No	No
Native Binary	No	No	No
netCDF	Yes	Yes, it was designed to be self-describing. CDL (schema-like) files are used to create files initially, but are not needed thereafter.	Not inherently. However, community-defined netCDF conventions can be used to write code that allows subsetting
BUFR/GRiB	No, you need tables to interpret the data	YES, but codes are used to describe projections, geophysical parameters, etc., so need to know these codes to interpret the data cronym List if needed	YES, but need the tables

Table 5.2.7 Data Standards Tools Support

Evaluation Questions			
Data Standard	Does the standard have freeware support?	Does the standard have COTS support?	
HDF	Yes, NCSA tools ImageMagick HDFLook HDFExplorer WIM H5View	Yes, PCI ENVI ER Mapper ERDAS-Imagine HDF Explorer IDL ImageMagick MATLAB Mathematica NCL Noesys PV-Wave WIM	
HDF-EOS	Yes, EOSView HE5View Webwinds	Slow: ENVI IDL MATLAB Noesys	
GeoTIFF	Open-source GIS tools (GRASS) and libraries for C (libgeotiff), Java (JAI), Python, etc.	Widespread: PCI-Geomatica RSI-ENVI ESRI-ArcView SoftDesk-AutoCAD ER Mapper ERDAS-Imagine Laser-Scan MapInfo MicroImages Intergraph-GeoMedia ENVI/IDL ¹	
Fast Format	GRASS, GDAL, OSSIM (Simple format, so manual import is common)	Moderate: PCI ENVI ER Mapper ERDAS-Imagine MicroImages (Simple format, so manual import is common)	
Native Binary	No	No	

_

¹ Any tool that reads a TIFF file should also read a GeoTIFF file (though most will complain about the extra "unsupported" tags). The packages listed here are those that use the additional information contained in a GeoTIFF file to geolocate the data.

Evaluation Questions		
Data Standard	Does the standard have freeware support?	Does the standard have COTS support?
netCDF	DODS GMT Linkwinds GrADS VisAD etc.	AVS Environmental workbench IDL interface IRIS Explorer Module MATLAB NCAR graphics Noesys PPLUS PV-Wave Silver Dicer WXP
BUFR/GRiB	Bufkit, GrADS	Limited

Table 5.2.8 Semantic Completeness

	Completeness Questions	
Data Standard	Can the format convey the data's precise spatial location?	Can the format convey the units of measure, the observation parameters (e.g., spectral bands), accuracy estimates (error bars), and other elements needed to understand the data?
HDF	Yes	Yes
HDF-EOS	Yes	Yes
GeoTIFF	Yes (basically)	No
Fast Format	Yes (basically)	No
Native Binary	N/A	N/A
netCDF	Yes	Yes
BUFR/GRiB	Yes with qualifications	Yes with qualifications

See Acronym List if needed

Based on the analysis of data standards using the eight criteria defined and the evaluation questions, each standard was giving a rating (low, medium and high) for each criterion. This rating is based on the answers to the evaluation questions. If a data standard can satisfy all the evaluation questions, a high rating is giving. If a data standard cannot satisfy any of the evaluation questions, a low rating is giving. If a data standard can satisfy some of the evaluation questions, a medium rating is giving. Table 5.2.9 summarizes the results.

Table 5.2.9 Data Standards Evaluation

	Data Packaging Standards								
Criteria	HDF	HDF- EOS	GeoTIFF	netCDF	Fast Format	BUFR/ GRiB	Native Binary		
Interoperability	High	<mark>High</mark>	High	High	<mark>High</mark>	Medium	Medium		
Availability	High	<mark>High</mark>	High	<mark>High</mark>	Low	Low	Medium		
Portability	High	High	High	High	High	Medium	Medium		
Evolvability	High	<mark>High</mark>	Medium	<mark>High</mark>	Low	Medium	Low		
Extensibility	High	Medium	High	High	Low	Medium	Low		
Self-describing	High	High	Medium	<mark>High</mark>	Low	Low	Low		
Tools Support	High	High	High	High	Medium	Medium	Low		
Completeness	<mark>High</mark>	High	Low	<mark>High</mark>	Low	Medium	N/A		

Table 5.2.9 shows that:

- HDF and netCDF receive high ratings for all eight criteria. This suggests that HDF and netCDF are good candidates as transmission or archive standards.
- Many standards, including HDF, HDF-EOS, GeoTIFF, and netCDF receive high ratings for Tools Support. This suggests that these standards can be used as analysis standards. Different user communities may prefer one standard over the others based on their familiarity with the standard and the simplicity and ease of use of the standards.

5.3 Metadata and Documentation Standards Evaluation

This section analyzes five metadata standards and one documentation standard used in heritage missions and related user communities:

- The Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata;
- ISO's draft standard on geographic metadata (ISO 19115);
- The EOSDIS (Earth Observing System Data and Information System)
 Core System Core Metadata Standard;
- The Directory Interchange Format (DIF) of the Global Change Master Directory (GCMD);
- Dublin Core Metadata Elements;
- The EOSDIS Data Gateway (or Information Management System (IMS) V0) Guide.

The above metadata and documentation standards are analyzed according to the same criteria used for data format standards. Table 5.3.1 shows the analysis results.

Table 5.3.1 Metadata and Documentation Standards Evaluation

			Documentatio				
	Metadata Format						
Criteria	ISO 19115	FGDC Content Standard	ECS Data Model	GCMD DIF	Dublin Core	Guide	
Interoperability	High: FGDC CSDGM, GCMD-DIF, etc. to be redefined as profiles of ISO-19115.	High: GILS, ECS, CIP (via GEO profile of Z39.50 API). To be reconciled w/ ISO 19115.	High: GCMD, FGDC compatible	High: ->ISO ->FGDC ->Dublin Core ->ANZLIC	High: FGDC, GILS, GCMD DIF	Medium: ->FGDC	
Availability	Low: draft on private ISO Website; final std. will cost	High: standard is on FGDC website	High: SDP and MDT Toolkits on HDF-EOS website	High: DIFs & DIF authoring tool available on web	High	High: Guide template available on web	
Portability	High content w/ xml schema	High: content w/ xml schema	High: Portable, can be created w/ text editor	High: can be created w/ text editor	High	High: can be created w/ text editor	
Extensibility	High (Built for profiles & extensions)	Moderate (extensions exist)	High: Can be extended	High: Can be extended	High: Can be extended	High: Can be extended	
Evolvability	High (many orgs. maintain it)	Moderate	Medium: Will be supported by ESDIS/EMD	High: GCMD group maintain the DIFs	Moderate	High: EDG sci/ops maintain the Guide	
Self-describing	Moderate	Moderate	Medium: Some attributes require documentation to understand	High: Collection level metadata self- describing, no granule metadata	Moderate	High: Guide template is self- describing	
Tools Support	Low: Emerging translators to/from FGDC; etc. No vendor support yet	High: fgdcmeta + xtme SMMS Metamanage r, MetaStar mp, cns Many vendor support	High: SDP toolkit Metadata works QAMUT EDG TerraWhom ECHO HDF (EOS) UFM No vendor support	High: Authoring tool DIF to XML Open access API Science keywords interface No vendor support	High: Many: DCMI Metadata Registry, DC-assist, Nordic DC metadata creator, DC-dot, Mantis, EdNA Metadata wizard	Medium: Converter Card creator Test bed Verifier No vendor support	

	Metadata Format					
Criteria	ISO 19115	FGDC Content Standard	ECS Data Model	GCMD DIF	Dublin Core	Guide
Completeness	High: carry semantic descriptive elements of the metadata	High: carry semantic descriptive elements of the metadata	High: carry semantic descriptive elements of the metadata	Low: Only semantics descriptive elements for collection metadata	High: carry semantic descriptive elements of the metadata	N/A (not applicable)

As shown in Table 5.3.1, all five metadata standards and the documentation standard receive similar ratings for most of the criteria used. 1This is because the metadata standards analyzed, FGDC CSDSM, ISO 19115, and ECS data model, are all based on each other. ISO 19115 was originally based on FGDC CSDSM version 1.0 and current FGDC "is consistent with the emerging ISO draft standards". FGDC will adopt ISO 19115 standard when it becomes final in 2002. The ECS data model is FGDC compliant, and FGDC Remote Sensing Extensions adopted many of the attributes in the ECS data model. Since ISO 19115 is an international metadata standard, it seems natural for FGDC and ECS data model to adopt ISO 19115 when it becomes final. ISO 19115 receives a low rating for "Availability" and "Tools Support" because the final ISO19115 standard has not been published.

GCMD is a metadata (collection level only) standard for on-line catalog access. Therefore, it is different from the ISO19115, FGDC CSDGM, and ECS data model. GCMD has been widely used in NASA Earth Sciences and GCMD DIF has been cross-mapped to ISO 19115, FGDC CSDSM, and Dublin Core elements. GCMD receives a low rating for "Completeness" because it has only semantic descriptive elements for collection metadata.

Guide document standard is interoperable with the FGDC metadata clearinghouse via Z39.50 API. However, Guide document standard is more suitable for Earth Science data sets.

5.4 Catalog Interface Standards Evaluation

This section analyzes five catalog interface standards used in heritage missions and related user communities:

- The EOSDIS Version 0 protocol;
- The GILS profile of the ANSI/ISO Z39.50 protocol;
- The GEO Profile of Z39.50 by the Federal Geographic Data Committee (FGDC);
- The CIP Profile of Z39.50 by the Committee on Earth Observation Satellites (CEOS);

The OGC Catalog Interface Specification by the OpenGIS Consortium;

The above catalog interface standards are analyzed according to the same criteria used for data and metadata standards. However, "Portability" is not used in the evaluation because interface standards specify communication procedures and structures between a client and server(s), which usually run on different platforms. So it is inherent that catalog interface standards are portable across different platforms. "Self-describing" is also removed from the evaluation because interface standards are protocol specifications between a client and server(s) and self-description does not apply to interface standards. Table 5.4.1 shows the analysis results.

Table 5.4.1 Catalog Interface Standards Evaluation

.	Catalog Interface Standards								
Criteria	EOSDIS V0	Z39.50 GILS	Z39.50 GEO	CIP	OGC				
iteroperability	Low: Merely tested for Interoperability with CIP	High: Interoperable with GEO, CIP & OGC	High: Interoperable with GILS, CIP, and OGC	High: Z39.50 portions are interoperable with GILS, GEO and OGC	High: Web profil is interoperable with GILS, GEO and CIP				
vailability	Moderate: Standard is controlled but available from NASA ESDIS	High: Standard can be downloaded from the website of Library of Congress	High: Standard can be accessed on www.blueangeltech.com	High: Standard can be downloaded From www.ceos.org	High: Standard is on www.opengis.or				
xtensibility	Can be extended using the EDG Data Dictionary	Can be extended using Z39.50 extended services	Can be extended using Z39.50 extended services	Can be extended using Z39.50 extended services	Can be extende using Z39.50 extended services				
volvability	Process exists to evolve the standard. NASA/GSFC ESDIS maintains the standard	Process exists to evolve the standard. Library of Congress maintains the standard	Process exists to evolve the standard. USGS maintains the standards	Process exists to evolve the standard. CEOS maintains the standards	Process exists to evolve the standard. Open GIS Consortium maintains the standards				
ools Support	Low: No commercial or free software tool support	High: Many freeware and commercial tools such as Isite MetaManager MetaStar	Moderate: Some freeware and commercial tools such as Isite MetaManager MetaStar	Moderate: Some freeware and commercial tools such as Isite MetaManager MetaStar	Moderate: Some freeware and commercial tool: such as Isite MetaManager MetaStar Earthscape WebAPI Service Manage				
ompleteness	Defines search, retrieval, order, and browse functions, information	Defines search and retrieval functions, information elements, and	Defines search and retrieval functions, information elements, and allowed values. Meets FGDC	Defines search, retrieval, order, and browse functions, information elements, and allowed values.	Defines search a retrieval functions, and optionally data access and				

0.46.45		Catalog Interface Standards					
Criteria	EOSDIS V0	Z39.50 GILS	Z39.50 GEO	CIP	OGC		
	elements, and allowed values. Meets NASA EOSDIS requirements.	allowed values. Meets the requirements of the library and museum communities	requirements.		management functions. Is data model neutral (no specification for information elements and allowed values)		
nplementation tatus	NASA EOSDIS uses the V0 standard in operation since 1994.	Widely used in federal and state library, museum, scientific, chemical, biological, and other communities since 1990's.	Used in several operational geospatial information systems, such as FGDC Clearinghouse, CEONet, MEL, since mid 1990s	Used in ESA INFEO system since 2001.	Several commercial vendors, such a GeoDan and ConTerra, have implemented OGC in their commercial products since the late 1990s.		

See Acronym List if needed

As shown in Table 5.4.1, the GILS, GEO, CIP profiles of Z39.50 and the OGC catalog standards receive similar ratings for most of the criteria used. This is mainly because GILS, GEO, CIP profiles and the OGC Web profile are all based on the Z39.50 protocol [2] that defines search and present operations. The search operation includes a guery and reguests the creation of a result set. The present operation returns subsets of a result set. Z39.50 is a stateful protocol. The base data model of Z39.50 is called bib-1 and is the main set of attributes appropriate to the bibliography world. GILS is developed on the basis of Z30.50 bib1 attributes, but in addition, contains extra attributes and references. GEO profile builds on bib-1 and GILS by extending the metadata model to include data elements from the Content Standard for Digital Geospatial Metadata maintained by the FGDC. CIP is both a profile and an extension of Z39.50. CIP is a profile of Z39.50 using messages, attributes, and elements in the standards as well as defining CIP-specific attributes and elements. CIP is an extension of Z39.50 through the addition of collections as a data structure and the addition of an ordering facility. Searches can be targeted at Collections, Product or Guide inventories. Retrieval of Browse products is supported. The OGC Catalog Interface standard has a general model and several profiles, e.g., CORBA profile, Web Profile. The OGC catalog interface Web profile is compliant to Z39.50. It contains constructs for the mandatory functions of search and retrieval, and for the optional functions of data access (order) using CIP order mechanism. The Z39.50 messages have been transformed into XML based messages for easier implementation. OGC catalog interface standard is data model independent [3]. CIP is aligned with GEO and GILS profiles [4].

The EOSDIS V0 system was originally developed as a prototype system for the EOSDIS. It gradually evolved into an operational system since 1994. Although it has low interoperability and low tool support, it is the only operational catalog interface standard used for inventory data by NASA Data Active Archive Center (DAAC) Alliance. It has recently been successfully tested for interoperability with the CIP profile.

5.5 User Surveys

In conjunction with the analyses of standards described above, we also conducted a user opinion survey on the data and metadata standards used in heritage and other missions to gain feedback from the user community. Not all of the criteria used in the standards analysis above were used in the survey because the survey was conducted before the criteria selection was refined. Also, there are several criteria used in the survey but not in the analysis. These criteria are more subjective than the more refined criteria used in the standards analysis and are defined below.

- Ease of use for producers.
- Ease of use for consumers.
- Acceptability Format is acceptable to a broad cross-section of likely users of the products.
- Suitability Has the proper descriptive power or precision for the task.
- Survivability The ability to be used by the community for many years.

There were a total of 45 survey respondents. Twenty surveys were returned from attendees of the NASA Science Data Processing Workshop, February 2002. Twenty-five surveys were collected from EOS User Working Group members at GSFC, LaRC, JPL, EDC, NSIDC, and ORNL DAACs (See Acronym List) and from other users. The survey results and some relevant statistics are summarized below.

5.5.1 Data Format Standards

Respondents were asked the question: "What weight should NASA give to the following criteria in evaluating a standard?" They were then asked to rate defined criteria with respect to "what weight NASA should apply when evaluating a standard," using a scale from one to six, where one was the lowest and six was the highest. Users gave "Ease of Use for Consumers" the highest rating (5.6) while "Evolvability" and "Ease of Use for Producers" received the lowest ratings. The statistics are summarized in Table 5.5.1.

Table 5.5.1 Survey Ratings of Attribute Importance

		Criteria							
Statistic	Ease of Use For Producer	Ease of Use For Consumer	•	Acceptability	Availability	Portability	Evolvability	Suitability	Interoperability
Average	4.3	5.6	5.3	5.1	5.1	5.4	4.6	4.6	4.8
Mode	4	6	6	5	6	6	4	5	6
Sample Size	41	41	42	42	41	42	41	39	41

Survey respondents were then asked to rate different data format standards using the set of criteria from the previous question. They were also ranked on a scale from one to six, where one was the lowest and six was the highest.

The survey of data standards used in the heritage missions indicated that users are most familiar with the HDF data standard. Thirty-five of the 45 total respondents were familiar with and rated HDF, while only one respondent rated Fast Format and only three rated BUFR. The sample sizes of the responses concerning Fast Format and BUFR were so small that these two formats were deleted from the results. Overall, the number of respondents is small (total respondents is 45, however, the sample size for a particular standard and particular criteria ranges between 5 and 35 as shown in Table 5.5.2); thus, drawing decisive conclusions from the survey is difficult. For example, approximately one-third of the people surveyed were not familiar with more than three data standards, and in many cases they gave the standard they were most familiar with the highest rating. However, the survey still shows some interesting findings.

- Binary format received the highest rating for most of the criteria used including Interoperability, Acceptability, Availability, Survivability, Evolvability, and Ease of Use. However, Binary was rated the lowest for Portability.
- HDF and netCDF were rated highest for Suitability and Portability.
 However, they were rated lowest for Ease of Use. This is due to the steep learning curve as indicated by the people surveyed.
- GeoTIFF was rated the second highest for Ease of Use. It was also rated high for Interoperability and Acceptability. However, it was rated the lowest for Suitability, Availability and Evolvability.

A summary of the statistics is shown in Table 5.5.2.

Table 5.5.2 Data Standards Survey Evaluation

	Type of		Data Format					
Criteria	Statistic	HDF	HDF- EOS	netCDF	GeoTIFF	Binary		
Interoperability	Average	4.3	3.6	3.6	4.5	4.9		
	Mode	6	4	3	5	5		
	Sample Size	28	24	10	8	12		
Acceptability	Average	4.4	3.9	4.1	4.3	4.8		
	Mode	6	3	4, 6	6	5, 6		
	Sample Size	34	27	15	9	13		
Availability	Average	4.7	4.2	4.9	3.4	5.1		
	Mode	6	6	6	3, 4	5		
	Sample Size	35	15	15	9	12		
Portability	Average	4.8	4.5	5.1	4.7	4.6		
	Mode	6	6	6	6	5		
	Sample Size	34	28	12	11	12		
Evolvability	Average	4.4	4.1	3.8	3.7	4.7		
	Mode	6	3, 6	6	4	6		
	Sample Size	20	20	5	6	9		
Suitability	Average	4.9	5.1	4.6	3.9	4.1		
	Mode	5	5	6	4	5		
	Sample Size	28	25	9	9	9		
Ease of use for	Average	3.5	3.6	3.6	4.8	5.0		
Consumer	Mode	4	3	2	4, 6	5		
	Sample Size	31	27	12	10	13		
Survivability	Average	4.7	4.5	4.7	4.7	5.5		
	Mode	5	4, 5	5 , 6	4	6		
	Sample Size	29	24	12	10	13		

Table 5.5.3 summarizes the results of four essay type questions from the opinion survey. An interesting result is that the HDF format received the most responses for all of the questions. About 60% of people surveyed listed that they had success with HDF and about 30% of the people listed HDF as a standard they foresee being used in the future. However, about 25% of the people surveyed listed that HDF is an impediment to their work. HDF-EOS, netCDF, and binary also received high ratings as a standard with which respondents had success (about 43% had success with HDF-EOS, 27% had success with Binary, and 23% had success with netCDF).

A detailed analysis of the results indicates that most of the respondents who had success with HDF are data producers from the computer/data manipulation areas (75% of them had success), atmospheric scientists (69%), and oceanographers (50%).

Respondents who had problems with HDF are atmospheric scientists (25% of them had problems), computer/data manipulation specialists (25%), and environmental scientists (20%). The most common complaint was that HDF was too complicated, making it difficult to learn (7 out of the 12 respondents), and that there was a lack of available tools (5 out of the 12 respondents). One atmospheric scientist seemed to capture this idea fairly well: "HDF, HDF-EOS, and netCDF are all initially more difficult, because they are more complex. But it isn't really an impediment as long as you have the right tools." Other impediments included huge file size, slow conversion from HDF4 to HDF5, and performing compression in HDF libraries. Similar complaints were listed for HDF-EOS, but also lack of support by RSI and cryptic failure messages were mentioned.

For other data formats, one atmospheric scientist said that GeoTIFF showed a dependency on machines. Another atmospheric scientist called BUFR, "ancient and primitive." The lone comment from an oceanographer for the GRiB format labeled it as difficult to use (worse than HDF). And one data producer said that binary was platform dependent and hard to verify.

Respondents who recommended HDF as a future standard are data producers (33% of them recommended), atmospheric scientists (25%), and 17% of oceanographers. Of the 21 respondents that commented on formats they foresee emerging in the future, eight (four atmospheric scientist, three data producers, and one environmental scientist) foresaw some form of HDF (HDF4, HDF5, or HDF-EOS) for the future because it is a powerful and versatile format. Four respondents (two data producers and two atmospheric scientists) felt some form of HDF was inevitable/mandated for the future.

Concerning other data formats, one oceanographer foresees and recommends GeoTIFF because it's easy to use, one atmospheric scientist says that ASCII is obvious for small data sets, and two respondents (an atmospheric scientist and an oceanographer) foresee/recommend netCDF because IDL supports it or because younger scientists are adopting it so it's growing in popularity. An oceanographer and an atmospheric scientist said that they foresee binary in the future because it has many tools and is widely available and because it is easy to use. Two scientists (atmospheric and environmental) recommend binary because it is simple, while one data producer recommends binary because many models require binary inputs.

Other comments on what the respondents foresee/recommend include: 1) atmospheric scientist – "Simple storage layer coupled with sophisticated connectivity layer. Let data producers have more freedom to use most appropriate format but force connectivity rules," 2) environmental scientist – "Something more easily accessible across platforms, software packages," 3) data producer – "Provide free, portable tools for ingesting and reformatting HDF and HDF-EOS. End-users will want to easily convert to formats such as ERDAS, ENVI/IDL, PCI, and Arcgrids."

Table 5.5.3 Summary of Survey Essay Questions

Question	Data Format						
Question	HDF	HDF-EOS	netCDF	GeoTIFF	Binary		
Data Format Used Successfully	27	19	10	2	12		
Data Format was Impediment	12	9	1	1	1		
Data Format Foresee Emerging	13	8	4	1	4		
Data Format Recommend	13	6	1	2	3		

5.5.2 Metadata Format Standards

For the metadata section, respondents of the user questionnaire were given a series of questions identical to those asked in the data format section of the survey. Of the 45 total surveys collected, only 16 people responded to the metadata questions. Thirteen of those were returned from attendees of the NASA Science Data Processing Workshop, February 2002. Three of those surveys were from EOS User Working Group members.

As in the data format portion of the survey, respondents were given a series of eight questions with a set of defined criteria as applied to particular metadata formats. They were asked to rate each metadata format (see headings in Table 5.5.4) with respect to the criteria using a scale from one to six, where one was the lowest and six was the highest.

The survey of metadata standards used in the heritage missions indicates that users are most familiar with the ECS Data Model, with as many as 11 responses out of the 16 total metadata responses, while only one respondent rated ISO and only two rated GCMD. The sample sizes of the responses concerning ISO and GCMD were so small that these two metadata formats were deleted from the results. Overall, the number of respondents is extremely small (as mentioned above) with the sample size for a particular standard and particular criteria ranging from 3 to 11 as shown in Table 5.5.4; thus, drawing decisive conclusions form the survey is difficult.

From the limited samples shown in Table 5.5.4, the FGDC content metadata and ECS data model received comparable ratings for most of the criteria. The FGDC content metadata received a little higher ratings than the ECS data model on Acceptability, Availability, Evolvability, and Survivability, while the ECS data model received a little higher ratings than the FGDC metadata model on Portability and Interoperability.

Table 5.5.4 Metadata Standards Survey Evaluation

Criteria	Statistics	Metadata	Format
Orneria	Otatistics	FGDC	ECS
Interoperability	Average	3.0	3.4
	Sample Size	3	7
Acceptability	Average	4.0	3.8
	Sample Size	5	10
Availability	Average	4.0	3.5
	Sample Size	4	11
Portability	Average	3.0	4.0
	Sample Size	4	10
Evolvability	Average	5.0	3.7
	Sample Size	2	9
Suitability	Average	4.4	4.5
	Sample Size	5	11
Ease of use for	Average	3.4	3.5
Consumer	Sample Size	5	10
Survivability	Average	5.6	4.4
	Sample Size	5	9

Table 5.5.5 summarizes the results of four essay type questions from the opinion survey. Sample sizes are very small as only 11 respondents answered the first question, 4 respondents answered the second and third questions, and 3 people answered the fourth question. The most significant observation is that 8 (73%) of the 11 respondents listed that they had success with the ECS Data Model. Only 27% claimed success using the FGDC Content Standard. 27% of the 11 respondents said that the ECS Data Model was an impediment, while 9% claimed the FGDC content standard was an impediment to their research. A detailed analysis of the results indicates that the respondents who claimed success with the ECS data model are data producers (5) and oceanographers (2). Some of the data producers (3) who claimed success with the ECS data model also indicated that they had problems with the ECS data model. Most of the complaints related to the ECS data model are that it is too complex, not flexible, not consistently applied, and that there is not enough tool support. Respondents who claimed success with the FGDC content standard include one data producer, one environmental scientist, and one atmospheric scientist. The environmental scientist who had used FGDC successfully also indicated that FGDC is an impediment in that there are too few tools and little portability, documentation, and interoperability. In terms of metadata that the respondents foresee in the future, one suggests adopting ISO 19115 to replace the current FGDC and adopting the FGDC extensions for remote sensing based on the ECS data model. One respondent recommends XML standard descriptions and defining XML DTD/schema for all the specific applications. One respondent

suggests refining the ECS data model, dropping most of the groups/classes, and attaching metadata to files.

Table 5.5.5 Summary of Metadata Survey Essay Questions

Question	Number of Times Metadata Format Was Listed			
	FGDC	ECS		
Data Format Used Successfully	3	7		
Data Format was Impediment	1	3		
Data Format Foresee Emerging	1	2		
Data Format Recommentd	1	2		

References:

[1] Jim Frew, 1998, How to Think about Data Formats, http://spso.gsfc.nasa.gov/diss/Meetings/19981005/frew_pres_9810.html.

[2] Information Retrieval (Z39.50): Application Service Definition and Protocol Specification, Z39.50 Maintenance Agency, 1995

3 OpenGIS – Catalog Interface Implementation Specification (version 1.0), OpenGIS Project Document 99-051s

[4] CIP Alignment with other Z39.50 profiles, 1997, European Commission

6.0 Summary

We have surveyed the standards for data and metadata that are in use in heritage missions or under consideration by the missions expected to be in formulation in the near-term. Lessons learned from heritage missions, some of the NOAA missions, and STDS have been reviewed. In particular, data and metadata standards in use in heritage missions and other EOS missions have been analyzed using a suite of criteria. Simple statistics and results from our user interview/survey to gather data producers' and data users' feedback on data and metadata standards are presented.

The highlights of lessons learned from heritage missions and other standards are summarized below.

- Multiple data distribution formats are used for some heritage missions, such as Landsat-7 and QuikSCAT/SeaWinds, to satisfy the diverse requirements from the user communities. Many NOAA POES missions use multiple data distributions formats to give users the flexibility to select the best data formats for their applications. This adds to the workload of NOAA agencies, such as NOAA NESDIS and NOAA NCDC, requiring them to develop and maintain different data translation tools in order to support different requirements from their user communities.
- Many different versions of HDF-EOS have been implemented for Terra data products, thus, creating problems for data interchange between mission instrument teams and users because different readers may be needed to read different implementations of the HDF-EOS data format. New EOS missions have realized this problem. For example, the Aura mission instrument teams have decided to adopt a uniform set of HDF-EOS file format guidelines so that data products from any Aura instrument are easily interchanged, i.e., the same set of tools and I/O routines can be used for all of the Aura data products.
- An important lesson learned from several missions, including Jason-1, SeaWiFS, SeaWinds, and ACRIM, is to provide good user support and experienced help desk for HDF-EOS implementation and usage. Many missions indicated that the "handholding" should not be underestimated.
- The Spatial Data Transfer Standard (SDTS), the national spatial data transfer mechanism for all U.S. Federal agencies, fell short of its ambitious goals and the marketplace was slow to accept and support it. The SDTS experience illustrates the importance of keeping pace with technology and market trends and emerging expectations, even after capturing initial requirements.

We devised eight standards criteria in order to objectively evaluate data and metadata standards. The results from the analysis of data format standards and metadata standards using the eight evaluation criteria are summarized below.

- HDF, and netCDF received high ratings for the evaluation criteria such as Interoperability, Availability, Evolvability, Portability, Extensibility, Tools Support, Completeness, and Self-describing. Many standards, including HDF, HDF-EOS, GeoTIFF, and netCDF received high ratings for Tools Support.
- BUFR/GriB and Fast Format in general were rated low to medium for many evaluation criteria, mainly Self-describing, Availability, and Completeness.
- Native binary received low ratings for Evolvability, Extensibility, Selfdescribing, and Tools Support (Chapter 5). Based on our analysis, native binary does not constitute a data standard and it cannot be compared alongside open consensus standards.
- The GILS, GEO, and CIP profiles of Z39.50 and the OGC catalog standards receive similar ratings for most of the criteria used. This is mainly because GILS, GEO, CIP profiles and the OGC Web profile are all based on the Z39.50 protocol. NASA EOS has used Z39.50 based standards for collection and document search. But for inventory search, EOSDIS Version 0 is the only operational catalog interface standard used by NASA EOS.
- The metadata and documentation standards analyzed received similar high to medium ratings for most of the evaluation criteria as many of the metadata standards, such as FGDC CSDGM, ISO 19115, and ECS data model, are all based on each other. We note that metadata standards are converging on the ISO 19115 when it becomes final in the near future.

We conducted a total of 45 interviews and surveys of data users from the EOS User Working Group members at different DAACs and of data producers/users from the 2002 NASA Science Data Processing Workshop. Although the sample size is not large, the interview/survey results illustrate feedback from data users and data producers on the data and metadata standards currently in use in NASA missions. Statistics and results are described in Chapter 5. A summary of the statistics/results is presented here.

- All of the users/producers answered questions related to data format standards. Only one-quarter of the users/producers answered questions related to metadata standards. This indicates that data producers and data users care more about (or are more familiar with or is more relevant to them) data format standards than metadata, and that many of them also have strong feelings about the data format standards.
- Users/producers are most familiar with the HDF data standard as 35 of the 45 total respondents were familiar with, and rated HDF, while only one respondent rated Fast Format and only three respondents rated BUFR.

- Many of those interviewed/surveyed are not familiar with multiple data formats, with only one-half of the respondents familiar with more than two data standards. In many cases they gave the standard they were most familiar with the highest rating. The results, therefore, may be biased.
- The interview/survey results show that HDF and netCDF were rated highest for Portability and Suitability and lowest for Ease of Use. However, respondents did not give high ratings to HDF and netCDF on Interoperability, Acceptability, Availability, and Evolvability. On the contrary, respondents rated binary format the highest for Interoperability, Acceptability, Availability, and Evolvability. This could possibly be because the producers'/users' understanding of these criteria are different from what we described in Chapter 5. This could also be because users are more familiar with Binary format and favor Binary format rather than the HDF or netCDF formats.
- The majority (60%) of respondents indicated that they had success with HDF and about one-third of the respondents recommend HDF as a future standard for NASA mainly because it is a powerful and versatile format. However, about one-quarter of the respondents also point out that HDF was an impediment to their work because HDF was too complicated, making it difficult to learn, and that there was a lack of available tools.
- For the metadata standards surveyed, respondents are most familiar with the ECS Data Model, with as many as 11 responses out of the 16 total metadata responses, while only one respondent rated ISO and only two respondents rated GCMD DIF. The ECS data model and the FGDC content metadata received comparable ratings for most of the criteria. This result is correspondent to the results derived from the standards analysis.
- For future metadata standards, some respondents recommend adoption
 of the ISO 19115 to replace the current FGDC and adoption of the FGDC
 extensions for remote sensing based on the ECS data model. Others
 recommend XML standard descriptions for metadata and refining the ECS
 data model.

7.0 Conclusions

Recommendations on data interface standards, data packaging standards, metadata standards, documentation standards, and associated activities for the near-term missions are summarized below. Acknowledging that there are several levels of requirements/guidelines, the following keywords are used to differentiate between them.

- o *must* This is mandatory.
- should This guideline is mandatory except where valid reasons exist to allow for its modification. Care should be taken in modifying or ignoring these guidelines.
- may This is a guideline which, while the NTMS group suggests it is worthwhile, is not mandatory.

7.1 Data Packaging Standards

The SEEDS Near-term Missions Study (NTMS) group recommends that data be packaged in an interchange format for storage and be available in multiple data distribution formats. The interchange format is for sharing data among the ESE data systems components (*i.e.*, data centers including PI-managed Mission Data Centers, "Backbone" Processing Centers, Science Data Centers, Application Data Centers, Multimission Data Centers, and similar centers or systems). Data distribution formats are for end-users.

NTMS recommends the following for packaging of near-term mission standard products.

7.1.1 Data Distribution Formats Recommendations

- 1. Data distribution facilities *must* enable packaging of standard data products in multiple distribution formats.
- 2. Distribution formats *must* emphasize end-user needs and convenience.

Rationale

At present, and for the near future, most applications and end-user practices are file-based. In such use, the format of the data files or the API used to read the data files are key to data access. Some communities have coalesced around particular file formats or access tools. For the greatest success in reaching multiple application and science disciplinary uses, ESE must support the preferences of these communities. These distribution-packaging choices (most simply understood as the formats in which data are sent to users) allow users to have access to data in one of several well-used formats. NTMS finds that several missions (including Landsat-7, QuikSCAT/SeaWinds, and many NOAA)

missions) have successfully employed multiple data distribution formats to satisfy the diverse requirements of their user communities (see Chapter 3).

All "Backbone" Processing Centers, Science Data Centers, and Multimission Data Centers *must* support a limited set of distribution packaging standards based on their end-users' requirements. The choice of distribution packaging *must* be made with the target community in mind and governed by applicability to task and the convenience of end-users.

NTMS does not have particular recommendations for distribution packaging standards. Based on our study, GeoTIFF format, WMO BUFR and GRID formats, Landsat Fast Format, and the API standards of netCDF and HDF/HDF-EOS and others are appropriate distribution packaging options.

7.1.2 Data Interchange Formats Recommendations

- 1. Interchange data sets *must* use a recognized packaging standard. The choice of standard *must* emphasize completeness and self-description.
- 2. Most of the near-term missions have indicated an interest in using HDF/HDF-EOS or netCDF. We agree that HDF/HDF-EOS or netCDF are appropriate choices as interchange data formats among ESE data system components (*i.e.*, data centers).
- Each appropriate ESE Near-term mission community must develop a profile
 of HDF/HDF-EOS or netCDF appropriate not only to the narrow needs of a
 particular mission but also to the wider needs of the allied community (See
 Chapter 4 for a discussion of profiles).
- 4. The development of each community's profile *must* be a process involving mission science teams, interested end-users, and experienced consultants. SEEDS NTMS has found that "community based" standards are more closely followed than standards imposed by outside forces.
- 5. Each community's interchange format profile *must* be as specific as possible to eliminate differences between data products and allow for the generation of simple data packaging tools.
- 6. If the HDF-EOS geolocation makes sense for a particular community, it *must* develop its interchange format profile based on this standard.
- 7. Each community *should* review other Near-term mission's community interchange format profiles and incorporate sections of overlap in their profile.
- 8. The interchange formats *may* be used as distribution formats.

Rationale

An interchange packaging standard among PI-managed mission-data centers and other ESE data systems components ensures that data are completely and correctly transferred. Use of standards for this interchange increases the flexibility of the ESE data systems. New components can join with the ESE data systems to provide data services without negotiating one-to-one interface

agreements with each potential provider. The effect of using standard packaging methods will result in decreasing the complexity of the ESE data systems as a whole while increasing potential for participation and novel use of NASA Earth systems science data sets. A community-involved process for approving or developing these data packaging standards ensures that the standards are appropriate and reliable. The choice of interchange packaging standards must consider completeness and correctness of representing data and emphasize the self-descriptiveness and long-term stability of the standards.

While HDF/HDF-EOS and netCDF, with a community-developed profile, are not the only possible candidates for an interchange standard, at this time, NTMS finds that they are the best choices based on our study (See Chapter 5). In fact, HDF/HDF-EOS is the most commonly used data packaging standard in the heritage missions. ESML (Earth Science Markup Language) has the potential to provide a level of data description to enable interchange packaging; however, NTMS finds this technology is not yet sufficiently mature to recommend. There are certainly other choices for standard interchange packaging as well.

NTMS finds that it is unlikely that certain data format standards will fill the needs of a center-to-center data interchange standard. For example, GeoTIFF is a good distribution standard for georeferenced imagery to end-users, but it is incapable of conveying the full metadata needed for the near-term missions or handling non-image data such as atmospheric profiles. BUFR and GRIB are WMO standards designed for dissemination of weather station data and for the output of numerical weather prediction models. BUFR and GRIB may be used as distribution formats for ocean or atmosphere data products used by weather prediction modelers, but they are not suitable as interchange standards as they lack self-describing power, tool support, and other criteria (see Chapter 5).

Finally, some near-term missions have heritage in, or are considering the use of, custom (*a.k.a.*, binary) data formats. We recommend that while it may be appropriate to use custom formats for internal mission science work, and certain communities may find them appropriate for distribution packaging, such formats are unlikely to be acceptable as an interchange standard.

7.2 Metadata Standards Recommendations

- Metadata standards are converging on the ISO 19115 standard, as FGDC will adopt the ISO 19115 after it becomes final in 2002. We recommend ISO19115 as the metadata standard for the near-term missions. However, since the ISO19115 is not finalized yet and the tool support for it has yet to be developed, we recommend the following implementation strategy.
- 2. Data systems for near-term missions must be compliant with the EOSDIS Clearing HOuse (ECHO) metadata model, which will be used for the advertising and distribution of data from "Backbone" Processing Centers, Science Data Centers, Multi-mission Data Centers, and PI-managed Missions Data Centers. The ECHO metadata model is an XML implementation and a superset of the ECS data model. It provides a capability to map metadata

- into various content-equivalent representations. With the single investment of ECHO interoperability, near-term missions will be insulated from most metadata standards changes and will benefit from new ECHO interfaces as they become available.
- The ECHO implementation organization should be tasked with monitoring, developing, and maintaining metadata mapping capability between ECHO holdings and emerging FGDC remote sensing profiles of the ISO 19115 standard.
- 4. We recommend continuing the use of the GCMD as a catalog (or collection) metadata standard for the near-term missions. We further recommend that the GCMD should coordinate with ECHO to implement seamless, automated interoperability of data products and services so that data centers do not need to prepare collection (or dataset) level metadata separately from inventory-level metadata.

7.3 Documentation Standards Recommendations

- 1. The Guide standard should be maintained, and a community-based process for incorporating the Guide into the ECHO data model should be developed. The EDG Guide data set documentation standard is successful and generally adequate for minimal description of standard data products. However, the division of metadata between the EOSDIS Earth Science Data Model and the EOSDIS Guide Document appears arbitrary and is, we believe, a hindrance to the effective, efficient, and accurate discovery and use of EOS data.
- 2. Algorithm Theoretical Basis Documents (ATBDs) are written by EOS scientists for every EOS instrument product. There are no detailed specifications for ATBDs, only a suggested outline that includes theoretical background, algorithm description, and validation plans. The ATBD is typically referenced in the Guide document. The ATBDs should be permanently accessible with a stable web address so that links to these documents in the Guide documents will remain valid
- 3. NASA should perform a detailed analysis of the emerging XML-based documentation standards in the social and library/archive sciences. The purpose of performing a detailed analysis of the emerging XML-based documentation standards in the social and library/archive sciences is to: 1) borrow from their XML-based data models which combine free-text descriptions and constrained element lists in a hierarchical, cross-referenced fashion; 2) maximize interoperability with information systems outside of Earth Sciences; and 3) benefit from the extensive work that has been done defining the documentation required for long-term preservation of knowledge about digital objects. We are specifically referring to the Data Documentation Initiative (DDI) in the Social Sciences, the Metadata Encoding and Transmission Standard (METS) maintained by the Library of Congress, and the Reference Model for an Open Archival Information System (OAIS)

developed by the Council of the Consultative Committee for Space Data Systems.

7.4 Catalog Interface Standards Recommendations

- 1. We recommend continuing the use of catalog interface standards for collection, inventory, and document search using the EOSDIS V0 protocol or Z39.50 based standards such as GEO or CIP.
- Web Service standards, such as OGC Web Mapping Service, Web Coverage Service, and Web Registry Service standards, will have an impact on catalog interface standards in the future. SEEDS should direct/track developments in the science and business communities.

7.5 Standard Evolution Process and Other Activities Recommendations

- 1. For Earth systems science systematic measurements including EOS data, weather data, atmospheric and oceanographic modeling data, and land use/land processes data, there are several data packaging or API standards that have similar purposes. ESE should invest resources in guiding the evolution of these data formats through their respective governing processes with the goal of harmonizing them toward seamless interoperability. We recommend that particular attention be focused on guiding the evolution of netCDF, HDF/HDF-EOS, geoTIFF and the WMO BUFR and GRIB formats
- 2. SEEDS should empanel a Standards and Interfaces Evolution Process Working Group for developing and executing a plan for evolution of interchange packaging standards over the life of data sets. The benefits of adopting additional interchange packaging standards need to be weighed against the increase in cost and complexity that will occur with the addition of each new packaging standard.
- Near-term missions must plan for evolution of end user requirements for packaging of mission science data (including data distribution packaging formats, data distribution system interface, and metadata) over the lifetime of the missions.
- 4. The evolution of the interchange packaging standards *must* keep pace with technology and market trends and emerging expectations (see lessons learned from SDTS in Chapter 3). SEEDS *should* adopt new technology as it develops. The number of interchange packaging standards *should* be limited and as closely related to each other as is practical.
- 5. SEEDS should coordinate respective activities to support the near-term missions such as interchange packaging standards maintenance and translation tools development/maintenance. This group should also provide interchange packaging standards user training and help desk support to educate producers/consumers/tool vendors.

6. The development of conversion software for data distribution formats *should* be a separately funded task and the responsibility for this development *should not* necessarily fall upon the mission science teams.

Acronym List

ACRIM Active Cavity Radiometer Irradiance Monitor

ACRIMSAT Active Cavity Radiometer Irradiance Monitor SATellite

ADEOS Advanced Earth Observing Satellite
ADS Archive and Distribution Segment

AGS Alaska Ground Station

AIRS Atmospheric Infrared Sounder
AIX IBM's UNIX Operating System

ALI Advanced Land Imager

ALT Dual-Frequency Radar Altimeter
AMI Active Microwave Instrument

AMSR Advanced Microwave Scanning Radiometer

AMSU Advanced Microwave Sounding Unit

ANZLIC Australia New Zealand Land Information Council

APAS Astrophysical, Planetary, and Atmospheric Sciences Department

API Application Platform Interface

APID Applications Package Identification

ASCI Accelerated Strategic Computing Initiative

ASPS AIRS Science Processing System

ASTER Advanced Spaceborne Thermal Emission And Reflection Radiometer

ATMOS Atmospheric Observations Satellite

ATMS Advanced Technology Microwave Sounder
AVHRR Advanced Very High Resolution Radiometer

AVISO Validation and Interpretation of Satellites Oceanographic data

AVS Advanced Visual Systems

BOREAS Boreal Ecosystem-Atmosphere Study

BSQ Band Sequential

BUFR Binary Universal Format For Representation [Of Data]

C3S Command Control & Communication Segment

CAP Cooperative Agreements Program
CARS Climate Analysis and Research

CASE Computer Aided Software Engineering

CBS Commission for Basic Systems

CCI Carbon Cycle Initiative

CCIWG carbon cycle interagency working group

CCS Climate Calibration Segment

CDF Common Data Format

CDHF Central Data Handling Facility

CDL Common Data Form Language (used by netCDF

CDMS Climate Data Management Segment

CDR Climate Data Record

CEOS Committee on Earth Observation Satellites
CERES Clouds and the Earth's Radiant Energy System

CI Catalog Interoperability

CIP Catalog Interoperability Protocol

CLAES Cryogenic Limb Array Etalon Spectrometry

CMIS Conical Microwave Imager/Sounder CMS Climate Mission Storage System

CNES Centre National D'etudes Spatiales (France)

CNIDR Clearinghouse for Networked Information Discovery and Retrieval

COTS Commercial Off-The-Shelf CPF Calibration Parameter File

CPOZ Compressed Ozone

CrIS Cross-Track Infrared Sounder

CRTT Calibrated Radiance and Temperature Tape
C-SAFS Central Standard Autonomous File System

CSDGM Content Standard for Digital Geospatial Metadata

CZCS Coastal Zone Color Scanner

DAAC Distributed Active Archive Center

DAO Data Assimilation Office

DC Dublin Core

DEM Digital Elevation Model

DFD Deutsches Fernerkundungsdatenzentrum (German Remote Sensing Data Center)

DFR Dual Frequency Radar

DIAL Data and Information Access Link
DIF Directory Interchange Format

DLL Dynamic Link Library
DLT Digital Linear Tape

DMF Data Models and Formats

DMSP Defense Meteorological Satellite Program

DoD Department of Defense

DODS Distributed Oceanographic Data System

DOE Department of Energy DOMSAT Domestic Satellite

DOQQ Digital Orthophoto Quarter Quads

DORIS Doppler Orbitography And Radiopositioning Integrated By Satellite

DOS Disk Operating System

D-PAF German Processing and Archiving Facility

DPR Dual-frequency Precipitation Radar

DPS Data Processing System
DRFP Draft Request for Proposal
DRG Digital Raster Graphics
DSP Directory Service Protocol
DSWG Data System Working Group

DTD Document Type Definition
EA Engineering Analysis
ECHO EOS Clearing House

ECMWF European Centre for Medium-Range Weather Forecasts

ECS EOSDIS Core System EDC EROS Data Center

EDG Earth Observing Systems (EOS) Data Gateway

EDOS EOS Data and Operations System

EDR Environmental Data Record
EMOS ECS Mission Operations System
ENSO El Niño/Southern Oscillation

ENVISAT Environmental Satellite
EO-1 Earth Orbiting Satellite #1
EOC EOS Operations Center
EOS Earth Observing System
EOSAT Earth Observation Satellite

EOSDIS Earth Observing System Data and Information System

EP Earth Probe

EPA Environmental Protection Agency
EPSG European Petroleum Survey Group
EROS Earth Resources Observation System

ERS Earth Resources Satellite

ERSDAC Earth Remote Sensing Data Analysis Center (Japan)

ERTS Earth Resource Technology Satellite (later renamed Landsat 1 (Land Saltellite?)

ESA European Space Agency
ESC Engineering Support Center

ESCAT ESA scatterometer

ESDIS Earth Science Data and Information System

ESDT Earth Sciences Data Type
ESE Earth Science Enterprise

ESIP Earth Science Information Partner

ESIPS EOSDIS Science Investigator-Led Processing System

ESRI Environmental Systems Research Institute (GIS software company)

ETM Enhanced Thematic Mapper

EUV Extreme Ultraviolet
FAST-L7A FAST-Landsat 7 Format
FDF Flight Dynamics Facility

FGDC Federal Geographic Data Committee

FIFE First ISLSCP Field Experiment

FIPS Federal Information Processing Standard

FMI Finnish Meteorological Institute

FTP File Transfer Protocol
FX File Transfer Subsystem

GAC Global-Area Coverage

GCCP Global Carbon Cycle Program
GCMD Global Change Master Directory

GCTE Global Change and Terrestrial Ecosystems
GDAAC Goddard Distributed Active Archive Center

GDAL Geospatial Data Abstraction Library

GDR Geophysical Data Record

GEO profile Geospatial Metadata Application

GeoTIFF Georeferenced Tagged Image File Format

GES Goddard's Earth Sciences

GHRC Global Hydrology Resource Center
GILS Global Information Locator Service
GIS Geographic Information System

GIVIT Granule Insert Validation and Inspection Tool

GLAS Geoscience Laser Altimeter System
GLOBEC Global Ocean Ecosystem Dynamics

GMT Greenwich Mean Time

GOES Geostationary Operational Environmental Satellites

GOFC Global Observation of Forest Cover GOME Global Ozone Monitoring Experiment GPM Global Precipitation Measurement

GRACE Gravity Recovery And Climate Experiment

GRASS Geographic Resources Analysis Support System (public domain software)

GRiB GRidded Binary

GRS-1 Generic Record Syntax

GSFC Goddard Space Flight Center
GUI Graphical User Interface
HAO High Altitude Observatory
HDF Hierarchical Data Format

HDF-EOS HDF Earth Observing System (EOS) format

HDSBUV High-Density Solar Backscatter Ultraviolet Instrument (SBUV)

HE4 HDF-EOS based on HDF version 4 HE5 HDF-EOS based on HDF version 5

HIRDLS High-Resolution Dynamics Limb Sounder
HRDLS High-Resolution Dynamics Limb Sounder
HRPT High Resolution Picture Transmission

HSB Humidity Sensor of Brazil

HTTPD Hyper Text Transfer Protocol Daemon

IAS Image Assessment System ICD Interface Control Document

ICESat Ice, Cloud, And Land Elevation Satellite

IDL Interactive Display Language
IDN International Directory Network

IDPS Interface Data Processing System **IETF** Internet Engineering Task Force

IFOV instantaneous field of view

IGBP International Geosphere And Biosphere Research Program

IGDR Interim Geophysical Data Record **IGS** International Ground Stations **IMS** Information Management System **IPD** Information Processing Division **IPO** Integrated Program Office

International Satellite Cloud Climatology Project **ISCCP**

ISLSCP International Satellite Land Surface Climatology Project

ISO Greek prefix "iso" as used by the International Organization for Standardization

ITSS Information Technology and Scientific Services

IWGDMGC Interagency Working Group on Data Management for Global Change

JAI Java Advanced Imaging **JEB** Java EOS Browser

JERS Japanese Earth Resources Satellite

JMR Jason Microwave Radiometer JPL Jet Propulsion Laboratory **KLM** NOAA K-, L-, M- system

Koninklijk Nederlands Meteorologisch Instituut (Netherlands) **KNMI**

LAC local-area coverage

LaRC Langley Research Center

Laboratory for Atmospheric and Space Physics at the University of Colorado **LASP**

70

Large-Scale Biosphere-Atmosphere Experiment LBA

LDCM Landsat Data Continuity Mission

LGN Landsat Ground Network **LGS** Landsat Ground Station LIS Lightning Imaging Sensor

LP Level Processor

LPDS Level 1 Product Distribution System **LPGS** Level 1 Product Generation System

LPS Landsat Processing System **LTER** Long-Term Ecological Research **MBLA** Multi-Beam Laser Altimeter **MCF** Metadata Configuration File

METI Ministry Of Economy Trade And Industry (Japan) **MFLOPS** Millions of Floating Point Operations per Second

MINC Medical Image netCDF

MISR Multi-Angle Imaging Spectro-Radiometer MIT Massachusetts Institute of Technology

MLE Maximum Likelihood Estimator

MLS Microwave Limb Sounder MOBY Marine Optical Buoy data MOC Missions Operations Center

MODAPS MODIS Adaptive Processing System

MODIS Moderate-Resolution Imaging Spectroradiometer MOPITT Measurements Of Pollution In The Troposphere

MSCD Mirror Scan Correction Data MSFC Marhsall Space Flight Center

MSS Multispectral Scanners

MTMGW Machine-To-Machine Search and Order Gateway

MUSE Multi-User Science Environment

NASA National Aeronautics and Space Administration
NASDA National Space Development Agency (Japan)
NCAR National Center for Atmospheric Research

NCDC National Climatic Data Center

NCEP National Centers for Environmental Prediction

NCL NCAR Command Language

NCSA National Center for Supercomputing Applications

NDVI Normalized Differential Vegetation Index

NESDIS NOAA/National Environmental Satellite, Data, and Information Service

netCDF Network Common Data Format

NIVR Netherlands's Agency for Aerospace Programs

NMC National Meteorological Center

NOAA National Oceanic and Atmospheric Administration

NODC NOAA National Oceanographic Data Center

NPOESS National Polar Orbiting Environmental Satellite System

NPP NPOESS Preparatory Project
NRL Naval Research Laboratory

NRT Near Real-Time
NSCAT NASA scatterometer

NSF National Science Foundation

NSIDC National Snow and Ice Data Center NSSDC National Space Science Data Center

NTMS Near-Term Missions Standards
NTMS Near-Term Missions Study
NWP Numerical Weather Product
NWS National Weather Service
ODL Object Description Language

OFL Off-Line

OGC OpenGIS Consortium

OMB Office of Management and Budget

OMI Ozone Mapping Instrument

OMPS Ozone Mapping and Profiler Suite

ORNL Oak Ridge National Labs

OSDPD Office of Satellite Data Processing and Distribution

OSDR Operational Sensor Date Record

OSF Observation Schedule File

OSSIM Open Source Software Image Map

OSTM Ocean Surface Topography Measurement
OTTER Oregon Transect Ecosystem Research

PCD Payload Correction Data
PCF Process Control File

PCI GIS software by PCI Geomatics

PDPS Planning and Data Processing System

PDR Product Delivery Record
PDS Precipitation Data System
PGE Product Generation Executive

PI Principal Investigator

PMEL Pacific Marine Environmental Laboratory

PNG Portable Network Graphics
PO.DAAC Physical Oceanography DAAC

POES Polar-Orbiting Environmental Satellite
POSC Petrotechnical Open Software Corporation

PP Pre-Processors

PPLUS graphics package by Plot Plus Graphics

PR Precipitation Radar

PSA Product Specific Attribute

PV-wave Visualization package by Visual Numerics, Inc. QAMUT Quality Assurance Metadata Update Tool

QuickScat Quick Scatterometer

QuikTOMS Quick TOMS

RBV Return-Beam Vidicons

RDBMS relational database management system

RDF Resource Description Framework

RDR Raw Data Record
RFP Request for Proposal
RGB Red, Green, Blue

RSI Research Systems, Inc. RSS Remote Sensing Systems

S4P Scalable Script-Based Science Processor

SA Science Analysis

SAA Satellite Active Archive

SAGE Stratospheric Aerosols and Gas Experiment
SAIC Science Applications International Corporation

SAN Storage Area Network
SAR Synthetic Aperture Radar

SASS Seasat-A scatterometer system

1/21/2003 72

SBUV Solar Backscatter Ultraviolet Instrument

SCF Science Computing Facility

SCIAMACHY SCanning Imaging Absorption SpectroMeter for Atmospheric ChartographY

SCLI Science data server Command-Line Interface

SDP Science Data Processing

SDPS SeaWifs Data Processing System

SDS Science Data Segment SDSRV Science Data Server

SDTS Spatial Data Transfer Standard

SeaPAC SeaWinds Processing and Analysis Center SeaWiFs Sea-viewing Wide Field-of-view Sensor

SEE Solar EUV Experiment

SEEDS Strategic Evolution of ESE science Data and information Systems

SERF Service Entry Resource Format
SFDU Standard Formatted Data Unit
SGDR Sensor Geophysical Data Record

SGI Silicon Graphics, Inc.

SGML Standard Generalized Markup Language

SGS Svalbard Norway Ground Station

SIGF Solar Irradiance Gap Filler SIM Spectral Irradiance Monitor

SIPS Science Investigator Processing System

SMI Standard Mapped Image

SMMR Scanning Multichannel Microwave Radiometer

SMMS Spatial Metadata Management System SNOE Student Nitrous Oxide Experiment

SOLSTICE Solar Stellar Irradiance Comparison Experiment

SORCE SOlar Radiation and Climate Experiment

SPOT Systeme Pour l'Observation De La Terre (France)

SPS Science Processing System
SQL Structured Query Language
SRM Subscription Request Manager
SRTM Shuttle Radar Topography Mission

SSALT Single-Frequency Solid-State Radar Altimeter

SSM/I Special Sensor Microwave/Imager

SUTRS Simple Unstructured Text Record Syntax

SWIR Short Wave Infrared TBD To Be Determined

TDI Transport Data Interface

TES Tropospheric Emission Spectrometer

TIFF Tagged Image File Format
TIM Total Irradiance Monitor

TIMED Thermosphere Ionosphere Mesosphere Energetics and Dynamics

1/21/2003 73

TIROS Television Infrared Observation Satellite

TM Thematic Mapper

TMI TRMM Microwave Imager

TMR TOPEX Microwave Radiometer
TOMS Total Ozone Mapping Spectrometer
TOPEX Topography (Ocean) Experiment
TOVS Tiros Operational Vertical Sounder
TRMM Tropical Rainfall Measurement Mission

TSDIS TRMM Science Data and Information System

TSIS Total Solar Irradiance Sensor

TSU TSDIS Science User

UARS Upper Atmospheric Research Satellite

UCSS UARS CDHF Software System

UDDI Universal Description, Discovery, and Integration

UFM User Friendly Metadata

UML Universal Modeling Language

USDA United States Department of Agriculture

USFS USDA Forest Service

USGCRP US Global Change Research Program

USGS United States Geological Survey
USMARC U.S. Machine Readable Cataloging
UTM Universal Transverse Mercator

VCL Vegetation Canopy Lidar VDC Visual Database Cookbook

VIIRS Visible/Infrared Imager/Radiometer Suite

VIRS Visible and Infrared Scanner
VNIR Visible And Near Infrared
W3C World Wide Web Consortium
WAIS Wide Area Information System

WGISS Working Group on Information Systems and Services

WHO World Health Organization WIM Windows Image Manager

WMO World Meteorological Organization

WNS Wind Scatterometer

WOCE World Ocean Circulation Experiment

WRS Worldwide Reference System

WXP Weather Processor

XDR eXternal Data Representation
XML eXtensible Markup Language
XPS XUV Photometer System

XSL eXtensible Stylesheet Language

XSLT XSL Transformations
ZMT Zonal Mean Tape

1/21/2003 74